MULTIMEDIA AND APPLICATIONS

DIRECTORATE OF DISTANCE EDUCATION

MBA

Paper 4.4.3

ALAGAPPA UNIVERSITY

Karaikudi - 630 003 Tamil Nadu

Reviewer	
Dr. M. Selvam	Dean, Faculty of Management Alagappa University, Karaikudi

Authors

Anirban Mukhopadhyay & Arup Chattopadhyay: Units (2.0-2.3, 3.0-3.2, 3.5) © Anirban Mukhopadhyay & Arup Chattopadhyay, 2013

Vikas Publishing House: Units (1, 2.4-2.9, 3.3-3.4, 3.6-3.9, 4, 5, 6) © Reserved, 2013

VIKAS® PUBLISHING HOUSE PVT. LTD.
E-28, Sector-8, Noida - 201301 (UP)
Phone: 0120-4078900 • Fax: 0120-4078999
Regd. Office: 576, Masjid Road, Jangpura, New Delhi 110 014
• Website: www.vikaspublishing.com • Email: helpline@vikaspublishing.com

Work Order No. AU/DDE/D2/Printing/31/2013-14 Date: 14.03.2013 Copies 1000

SYLLABI-BOOK MAPPING TABLE

Multimedia and Applications

Syllabi	Mapping in Book
UNIT 1 **Multimedia in Use and Technology:** Introducing Multimedia —Multimedia Definition—Need, Benefits and Problems—System Components—Multimedia Platforms—Development Tools: Types—Cross Platform Compatibility—Commercial Tools—Standards.	**Unit 1:** Multimedia in Use and Technology **(Pages 3-62)**
UNIT 2 **Media Types:** Non-Temporal—Text, Image, Graphics—Temporal—Analog, Digital Audio/Video, Music, Animation, Other Media Types—Extended Images, Digital Ink, Speech, Audio.	**Unit 2:** Media Types **(Pages 63-118)**
UNIT 3 **Digital Video and Image Compression:** Evaluating a Compression System—Redundancy and Visibility—Video Compression Techniques—Image Compression Standards—JPEG, MPEG, DV1.	**Unit 3:** Digital Video and Image Compression **(Pages 119-144)**
UNIT 4 **Object Oriented Multimedia:** Objects, Classes and Related Items—Multimedia Frameworks: Overview, Media Classes, Transform Classes, Format Classes, Component Classes.	**Unit 4:** Object Oriented Multimedia **(Pages 145-176)**
UNIT 5 **Multimedia Environments:** The CD Family, CD-i—Overview—Media Types—Media Organization—Architecture and Operations, Applications: Media in Real World—Multimedia on Networks—Training and Education.	**Unit 5:** Multimedia Environments **(Pages 177-220)**
UNIT 6 **Virtual Reality:** Intelligent Multimedia System—Desktop Virtual Reality (VR)—VR Operating System—Virtual Environment Display and Orientation Making Visually Coupled System Requirements—Intelligent VR Software Systems.	**Unit 6:** Virtual Reality **(Pages 221-243)**

CONTENTS

INTRODUCTION

NOTES

Computers have brought about major changes in all spheres of life and in the way we communicate. Today, it is extremely difficult to imagine a world without computers. The development of multimedia and computer graphics has made computers easier to interact with and enhanced for interpreting many types of data. Developments in computerized multimedia graphics have a profound impact on many types of media and have also revolutionized animation, movies and the video game industry.

In today's world, multimedia is a vital component of communication, entertainment and education. As a general term, multimedia, which literally means 'many media', can refer to the sharing of information utilizing at least two of our five sensory organs. Multimedia has its application in various areas including advertisements, art, education, entertainment, engineering, medicine, mathematics, business, scientific research and spatial temporal applications. This book will discuss multimedia in a digital context, i.e., digital multimedia. Digital multimedia is mostly computer generated and comprises text, images, audio, video and animation, and the hardware and software requirements for producing digital multimedia.

Technically, multimedia signifies a combination of various content forms, such as text, images, audio, animation, video and interactivity. Broadly speaking, multimedia can be categorized as linear and nonlinear. Linear multimedia is a progressive presentation of content, without random navigation facility or non-interactive, for example cinema presentation. Nonlinear multimedia allows interactivity, i.e., the user can control the progress of the content for example computer games. Computer animation is the art of creating moving images using computers. It is a subfield of computer graphics and animation. Multiple methods of achieving animation exist; the rudimentary form is based on the creation and editing of keyframes, each storing a value at a given time, per attribute to be animated. The 2-D/3-D graphics software interpolates between keyframes, creating an editable curve of a value mapped over time, resulting in animation. Multimedia, thus, includes a combination of text, audio, still images, animation, video and interactivity content forms. This book introduce you the exciting multimedia technologies, the significance of which is greatly felt in our day-to-day life.

This book, *Multimedia and Applications*, is aimed at giving the students a fair thought of the multimedia and its significant applications in today's world. The book follows the self-instruction mode or SIM format wherein each unit begins with an 'Introduction' to the topic of the unit followed by an outline of the 'Unit Objectives'. The detailed content is then presented in a simple and structured form interspersed with 'Check Your Progress' questions to facilitate a better understanding of the topics discussed. The 'Key Terms' are given on respective pages to help the student revise what he/she has learnt. A 'Summary' along with a set of 'Questions and Exercises' is also provided at the end of each unit for effective recapitulation.

UNIT 1 MULTIMEDIA IN USE AND TECHNOLOGY

NOTES

Structure

1.0 INTRODUCTION

In this unit, you will learn about multimedia and its related technology. Multimedia refers to a mixture of interactive media or data types, predominantly text, graphics, audio and video that are simultaneously delivered by a computer. Actually a multimedia system implies a combination of specified hardware components with certain minimum capabilities and compatible software that has an interactive interface studded with different media elements. The application areas for digital multimedia are continuously increasing since the last decade. It has become an indispensable tool for visualization and storage of the happenings around us.

You will learn about various system components. The most prominent part in a personal computer is the display system that is responsible for graphic display. The display system may be attached with a PC to display character, picture and video output. You will also learn about various multimedia devices like capture devices, storage devices, communication network devices, computer systems and display devices. The presentation devices are associated with the multimedia applications in GUI (Graphical User Interface) environment. GUI allows users to select the available resources visually and control the presentation devices. It controls the virtual room lighting, hotspots, loudspeaker, printers, cameras, etc. The presentation device connects the computer using various wireless technologies, such as Bluetooth connection to produce the presentation online.

You will also learn about multimedia platforms. A true multimedia platform integrates and combines various multimedia devices and components.

Finally, you will learn about various development tools and multimedia standards. The key to successful multimedia production is a seamless integration of multimedia elements for graphic design, content management, production and packaging. The whole process of developing a multimedia package is called authoring. An authoring system is a collection of software tools that help in various aspects of multimedia production. In computer technology, cross platform or multi-platform refers to the unique characteristic of computer software which enables it in implementing methodologies for inter-operating on several computer platforms. The term 'Standards' refers to the specifications made by the systematic efforts approved by official standardization federations committed to the issue and can sometimes be termed as *official standards*. In some specific cases it is termed as *de facto standards* when it is widely accepted by the industry and/or the public.

1.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Learn about multimedia needs and benefits
- Discuss the various system components of multimedia
- Explain the significance of multimedia platforms
- Describe various types of developments tools and multimedia standards

1.2 INTRODUCING MULTIMEDIA

Multimedia: A mixture of interactive media or data types, predominantly text, graphics, audio and video that are simultaneously delivered by a computer

Multimedia refers to a mixture of interactive media or data types, predominantly text, graphics, audio and video that are simultaneously delivered by a computer. Actually a multimedia system implies a combination of specified hardware components with certain minimum capabilities and compatible software that has an interactive interface studded with different media elements. The difference between a multimedia system and a television (where you also get simultaneous presentation of multiple media) is that in television, the delivery is not interactive and the user cannot control the way things happen, whereas a PC-Multimedia system allows the user to control the elements that are delivered with a varied degree of navigational freedom through linked elements.

Though basically multimedia presentation is dependent on the processing power and data-storage capacity of the computer, some basic hardware (H/W) components that make a complete multimedia system are, as follows:

1. Devices like keyboard, mouse, joystick, touch screen by which the user can interact with the system.
2. A high-resolution screen and graphics accelerator card that can provide good quality still images, video clips; animations; text and graphics.
3. Speakers for speech and music output.
4. Microphone for audio recording.

5. Sound card and video grabber card to capture, digitize and edit audio and video material.
6. **CD-ROM drive** to play back pre-recorded source (/) material. CD-ROM drive is a device that can read information from a CD-ROM

With the rapid progress of the H/W industry, new generation processors and memory chips, various add-ons, computing accessories and devices upgraded kits are being continuously evolved. The PCs are becoming more and more powerful endowed with stunning multimedia capabilities. However, there is an international standard specification of a Multimedia PC (MPC) by an industry consortium called Multimedia PC Working Group (MPCWG), formerly known as Multimedia PC Marketing Council. The MPC specifications, which are upgraded from time to time, clearly define what should be the minimum features that a multimedia PC should have.

According to the latest MPC level-3 specification (the previous two being level-1 and level-2) a multimedia PC should have the following as the absolute minimum (even though a higher specification is recommended):

1. 75 MHz Intel Pentium Processor or equivalent.
2. 8 MB RAM.
3. 540 MB Hard Disk Drive with 15 ms access time, 1.5 MB/sec sustained throughout.
4. CD-ROM Drive with 250 ms access time, 600 KB/sec transfer rate.
5. 3.5", 1.44 MB Floppy Disk Drive.
6. Color Monitor with display resolution of 640 × 480 with 65,536 (16-bit) colors.
7. Video Playback (Full Motion Video): MPEG1 decoding support with its output being able to drive atleast a 320 × 240 pixel video window, at 30 fps and with 15 bits/pixel.
8. 101 key IBM compatible keyboard or equivalent.
9. Two button Mouse.
10. Audio board.
11. MIDI input-output port.
12. Joystick port.
13. Headphones and/or speakers.
14. Serial and Parallel ports.
15. Software: DOS 6.0 or later, Windows 3.11 or later.

This MPC level-3 specification was released in 1995. However, with respect to the current market standard MPC level-3 may seem to be quite outdated. After Windows 3.11, Windows 95, 98, 2000, and XP have come up with improved GUI and support a variety of multimedia devices. Moreover, 3GHz Pentium - 4 PC's with 80 GB HDD are now commonly used. In addition, most of the multimedia applications today require at least 256 MB RAM, if not more.

The multimedia devices and drivers are managed by the [mci] and [drivers] section of the Windows SYSTEM.INI file, that can be added and deleted using the Multimedia Properties control panel. The device properties can also be adjusted from there.

NOTES

CD-ROM drive: A device that can read information from a CD-ROM

By reading the SYSTEM.INI text file when it starts up, Windows knows what multimedia devices are present in the system and initializes them.

1.2.1 Need of Multimedia

NOTES

The application areas for digital multimedia are continuously increasing since the last decade. It has become an indispensable tool for visualization and storage of the happenings around us. In fact, digital multimedia can be effectively used wherever there is scope for representing some information interactively using text, sound, image, video, etc. The several key areas where multimedia technology has been effectively utilized for the benefits of the people are:

- ***Education and Training*****:** Digital multimedia libraries of interactive e-learning courses are available for individual as well as corporate-level training. It is a great boon of multimedia technology and is spreading rapidly all over the world as a very popular and dependable means of distance education. A quality e-learning module facilitates self-paced training with quality courseware prepared by competent and experienced faculty members that provides great relief and help to students who are otherwise deprived of classroom teaching. Even when classroom training is available, e-learning may effectively supplement such formal training with features, such as interactivity, multimedia demonstration, quizzes, etc. Further, e-learning is usually implemented through a web-based Learning Management System (LMS) that keeps track of the student's progress and performance.
- ***Entertainment Industry*****:** Any activity that gives pleasure to the audience is entertainment. It can be a show or performance by an individual or a group, such as a magic show, a theatre, a football match or a movie show. The industry that provides entertainment is called the entertainment industry. With the advent of digital multimedia, the entertainment industry has been using digital multimedia tools and techniques for creating special effects, developing interactive computer games, edit movies, create animation films as well as restore and enhance classic films of yesteryears. The audience may participate actively in a computer game or passively as in watching an animation film or a movie with special effects created using digital multimedia tools, but the role of multimedia is always there.
- ***Household Services*****:** Today, you can routinely use multimedia technology in various forms in your home, such as for self-paced education, home shopping (railway or airlines booking, product demonstration, e-filling of forms), communicating with distant relatives through multimedia chatting. Other application areas are video on demand, interactive TV, etc., that are gradually becoming popular.
- ***Business Services*****:** Multimedia technology is routinely applied in videoconferencing in a cost effective way, e-mail and multimedia chatting for routine official correspondence and transaction of important multimedia documents across the globe.
- ***Science and Technology*****:** Visualization and simulation can be done using multimedia technology for all branches of science and technology.
- ***Medicine*****: Interactive multimedia technology** can be used in implementing telemedicine projects to cure patients at far-off places. Multimedia databases provide support for queries related to medical science and patient related queries including case history, X-rays, scanned images, assessments, response, etc.

Interactive multimedia technology: It can be used in implementing telemedicine projects to cure patients at far-off places

1.2.2 Benefits and Limitations of Multimedia

The following are some of the benefits and limitations of multimedia:

Benefits

NOTES

- The portability of digital multimedia formats facilitates easy transportation and helps in manipulating information comprehensively.
- Careful use of media files can make a Web page, blog post or a report more engaging than simple plain text, emphasizes key pieces of information.
- It increases learning effectiveness in education and training programs.
- It is more appealing over traditional lecture-based learning methods.
- Offers significant potential in improving personal communications, education and training efforts.

Limitations

- Multimedia formats and the devices that play or store them require a constant supply of power and frequent updating that can be problematic in more remote areas.
- As technology rapidly evolves, compatibility between different devices can also be a problem when trying to move or play multimedia content.
- The factor of expense is unable to keep up with technology for financial reasons or geographic isolation.
- Adding multimedia increases the number of codec and plug-ins a browser needs to load the page which leads to slower loading times.
- Multimedia also leads to a third-party problem, such as if a video being removed from the referenced Website or links then it will leave a blank space in any post in which the video have been embedded.

1.3 SYSTEM COMPONENTS

In this section, you will learn about the display and input devices that are generally used to display objects.

Display Devices

The most prominent part in a personal computer is the display system that is responsible for graphic display. The display system may be attached with a PC to display character, picture and video output. Some of the common types of display system available in the market are,

1. Raster Scan Displays.
2. Random Scan Displays.
3. Direct View Storage Tube.
4. Flat Panel Displays.
5. Three Dimensional Viewing Devices.
6. Stereoscopic and Virtual Reality System.

NOTES

Pixel: As the smallest size object or color spot that can be displayed and addressed on a monitor

These display systems are often referred to as Video Monitor or Visual Display Unit (VDU). The most common video monitor that normally comes with a PC is the Raster scan type. However, every display system has three basic parts – the display adapter that creates and holds the image information, the monitor which displays that information and the cable that carries the image data between the display adapter and the monitor.

Before the major display systems are discussed, let us first know about some basic terms.

1. Pixel

A ***pixel*** may be defined as the smallest size object or color spot that can be displayed and addressed on a monitor. Any image that is displayed on the monitor is made up of thousands of such small pixels (also known as picture elements). The closely spaced pixels divide the image area into a compact and uniform two-dimensional grid of pixel lines and columns. Each pixel has a particular color and brightness value. Though the size of a pixel depends mostly on the size of the electron beam within the CRT, they are too fine and close to each other to be perceptible by the human eye. The finer the pixels the more the number of pixels displayable on a monitor-screen. However, it should be remembered that the number of pixels in an image is fixed by the program that creates the image and not by the hardware that displays it.

2. Resolution

There are two distinctly different terms, which are often confused. One is *Image Resolution* and the other is *Screen Resolution*. Strictly speaking, image resolution refers to the pixel spacing, i.e., the distance between one pixel and the next pixel. A typical PC monitor displays screen images with a resolution somewhere between 25 pixels per inch and 80 pixels per inch (ppi). In other words, resolution of an image refers to the total number of pixels along the entire height and width of the image. For example, a full-screen image with resolution 800 × 600 means that there are 800 columns of pixels, each column comprising of 600 pixels, i.e., a total of 800 × 600 = 4,80,000 pixels in the image area.

The internal surface of the monitor screen is coated with red, green and blue phosphors that glows when they are struck by a stream of electrons. This phosphor-coated material is arranged in an array of millions of tiny cells—red, green and blue, usually called dots. The *dot pitch* refer to the distance between adjacent sets (triads) of red, green and blue dots. This is also considered to be the same as the shortest distance between any two dots of the same color, i.e., from red-to-red or green-to-green like that. Usually monitors are available with a dot pitch specification 0.25 mm to 0.40 mm. Each dot glow with a single pure color (red, green or blue) and each glowing triad appears to our eye as a small spot of color (a mixture of red, green and blue). Depending on the intensity of the red, green and blue, different colors results in different triads. The dot pitch of the monitor thus, indicates how fine can be the colored spots that make up the picture, though electron beam diameter is an important factor in determining the spot size.

So, you understand that the pixel is the smallest element of a displayed image and that dots (red, green and blue) are the smallest elements of a display surface (monitor screen). The dot pitch is the measure of the screen resolution. The smaller the dot pitch, the higher the resolution, sharpness and detail of the image displayed.

NOTES

To use different resolutions on a monitor, the monitor must support automatic changing of resolution modes. Originally, monitors were fixed at a particular resolution, but for most monitors today, the display resolution can be changed using software control. This lets you use higher or lower resolution depending on the need of your application. A higher resolution display allows you to see more information on your screen at a time and is particularly useful for operating systems, such as Windows. However, the resolution of an image you see is a function of what the video card outputs and what the monitor is capable of displaying. To see a high resolution image, such as 1280 × 1024 requires both a video card capable of producing an image this large and a monitor capable of displaying it.

3. Image Resolution versus Dot Pitch

If the image resolution is more as compared to the inherent resolution of the display device, then the quality of the displayed image gets reduced. As the image has to fit in the limited resolution of the monitor, the screen pixels (comprising a red, a green and a blue dot) show the average color and brightness of several adjacent image pixels. Only when the two resolutions match, the image is displayed perfectly and only then the monitor is considered to be used to its maximum capacity.

4. Aspect Ratio

The aspect ratio of the image is the ratio of the number of X pixels to the number of Y pixels. The standard aspect ratio for PCs is 4:3; some resolutions even use a ratio of 5:4. Monitors are calibrated to this standard so that when you draw a circle it appears to be a circle and not an ellipse. Displaying an image that uses an aspect ratio of 5:4 will cause the image to appear somewhat distorted. The only mainstream resolution that uses 5:4 is the high-resolution 1280 × 1024. Table 1.1 shows the resolution their respective number of pixels and their aspect ratios.

Table 1.1 *Common Resolutions, Number of Pixels and Standard Aspect Ratios*

Resolution	Number of Pixels	Aspect Ratio
320 × 200	64,000	8:5
640 × 480	307,200	4:3
800 × 600	480,000	4:3
1024 × 768	786,432	4:3
1280 × 1024	1,310,720	5:4
1600 × 1200	1,920,000	4:3

5. Raster Scan Display

This type of display basically employs a Cathode Ray Tube (CRT) or LCD (Liquid Crystal Display) panel for display. The CRT works just like the picture tube of a television set. Its viewing surface is coated with a layer of arrayed phosphor dots. At the back of the CRT is a set of electron guns (cathodes) which produce controlled streams of electrons (electron beams). The phosphor material emits light when struck by these high-energy electrons. The frequency and intensity of the light emitted depends on the type of phosphor material used and energy of the electrons. To produce a picture on the screen, these directed electron beams start at the top of the screen and scan rapidly from left to right along the row of phosphor dots. They return to the leftmost position one line down and scan again, and repeat this to cover the entire screen. The return of the beam direction to the leftmost position one line down is

NOTES

called **horizontal retrace** during which the electron flow is shut off. In performing this scanning or sweeping type motion, the electron guns are controlled by the video data stream coming into the monitor from the video card, which varies the intensity of the electron beam at each position on the screen. The instantaneous control of the intensity of the electron beam at each dot is what controls the color and brightness of each pixel on the screen. All these happen extremely quickly, and the entire screen is drawn in a fraction (say, 1/60th) of a second.

An image in raster scan display is basically composed of a set of dots and lines; lines are displayed by making those dots bright (with the desired color) which lie as close as possible to the shortest path between the endpoints of a line.

6. Refresh Rate and Interlacing

When a dot of phosphor material is struck by the electron beam, it glows for a fraction of a second and then fades. As the brightness of the dots begins to reduce, the screen image becomes unstable and gradually fades out.

To maintain a stable image, the electron beam sweeps the entire surface of the screen and then returns to redraw it a number of times per second. This process is called **refreshing** the screen. After scanning all the pixel-rows of the display surface, the electron beam reaches the rightmost position in the bottommost pixel line. The electron flow is then switched off and the vertical deflection mechanism steers the beam to the top left position to start another cycle of scanning. This diagonal movement of the beam direction across the display surface is known as *vertical retrace*. If the electron beam takes too long to return and redraw a pixel, the pixel will begin to fade; it will return to full brightness only when redrawn. Over the full surface of the screen, this becomes visible as a *flicker* in the image, which can be distracting and hard on the eyes.

In order to avoid flicker, the screen image must be redrawn sufficiently quickly that the eye cannot tell that the refresh is going on. The *refresh rate* is the number of times per second that the screen is refreshed. The unit of frequency is measured in Hertz (Hz). The refresh rates are somewhat standardized; common values are 56, 60, 65, 70, 72, 75, 80, 85, 90, 95, 100, 110 and 120 Hz. Basically, higher refresh rates are preferred while viewing the monitor as it provides bright image on the screen, but the maximum refresh rate possible depends on the resolution of the image. The maximum refresh rate that a higher resolution image can support is less than that supported by a lower resolution image, because the monitor has more number of pixels to cover with each sweep. Actually, the support for a given refresh rate requires two things: a video card capable of producing video images that many times per second, and a monitor capable of handling and displaying the same number of signals per second.

Every monitor as a part of its specification should include a list of resolutions it supports and the maximum refresh rate for each resolution. Many video cards now include setup utilities that are pre-programmed with information about different monitors. When you select a monitor, the video card automatically adjusts the resolutions and their respective allowable refresh rates. Windows 95 and later versions extends this facility by supporting Plug and Play for monitors; you plug the monitor in and Windows will detect it, set the correct display type and automatically choose the optimal refresh rate.

Some monitors use a technique called *interlacing* to cheat a bit and allow themselves to display at a higher resolution than is otherwise possible. Instead of refreshing every line of the screen, when in an interlaced mode the electron guns sweep

alternate lines on each pass. In the first pass, odd-numbered lines are refreshed and in the second pass, even-numbered lines are refreshed. This allows the refresh rate to be doubled because only half the screen is redrawn at a time. The usual refresh rate for interlaced operation is 87 Hz, which corresponds to 43.5 Hz of 'real' refresh in half-screen interlacing. Figure 1.1 shows a schematic diagram of an interlaced raster scan.

NOTES

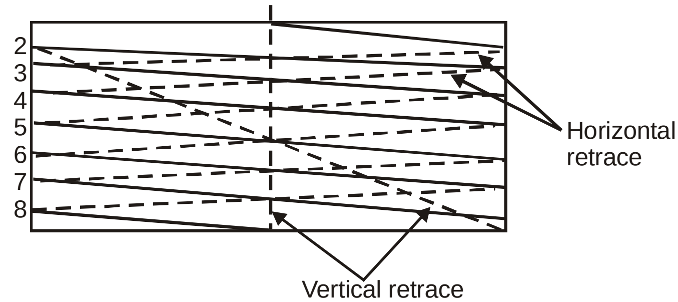


Fig. 1.1 *Schematic Diagram of Interlaced Raster Scan*

Figure 1.1, the odd-numbered lines represent the scanning of one half of the screen and the even-numbered lines represents the scanning of the other half. There are two separate sets of horizontal and vertical retraces.

Cathode Ray Tube or CRT

A **CRT** is a vacuum tube containing an electron gun and a fluores cent screen to view images. It is similar to a big vacuum glass bottle. It contains three electron guns that radiates a focussed beam of electrons, deflection apparatus (magnetic or electrostatic), which deflects these beams both up and down and sidewise and a phosphor-coated screen upon which these beams impinge. The vacuum is necessary to let those electron beams travel across the tube without running into the air molecules that could either absorb or scatter them.

CRT: A vacuum Tube containing a electron gun and a thuorexent screen to view images.

The primary component in an electron gun is a cathode (negatively charged) that is encapsulated by a metal cylinder known as the control grid. A heating element inside the cathode causes the cathode to be heated up as current is passed. As a result electrons 'boil-off' from the hot cathode surface. These electrons are accelerated towards the CRT screen by a high positive voltage applied near the screen or by an accelerating anode. If allowed to continue uninterrupted, the naturally diverging electrons would simply flood the entire screen. The cloud of electrons is forced to converge to a small spot as it touches the CRT screen by a focussing system using an electrostatic or magnetic field. Just as an optical lens focuses a beam of light at a particular focal distance, a positively charged metal cylinder focuses the electron beam passing through it on the center of the CRT screen. A pair of magnetic deflection coils mounted outside the CRT envelope deflects the concentrated electron beam to converge at different points on the screen in the process of scanning. Horizontal deflection is obtained by one pair of coils and vertical deflection by the other pair. The deflection amount is controlled by adjusting the current passing through the coils. When the electron beam is deflected away from the center of the screen, the point of convergence tends to fall behind the screen resulting in blurred (defocussed) display near the screen edges. In high-end display devices, this problem is eliminated by a mechanism which dynamically adjusts the beam focus at different points on the screen. Figure 1.2 shows the diagram of a CRT raster scan display device.

NOTES

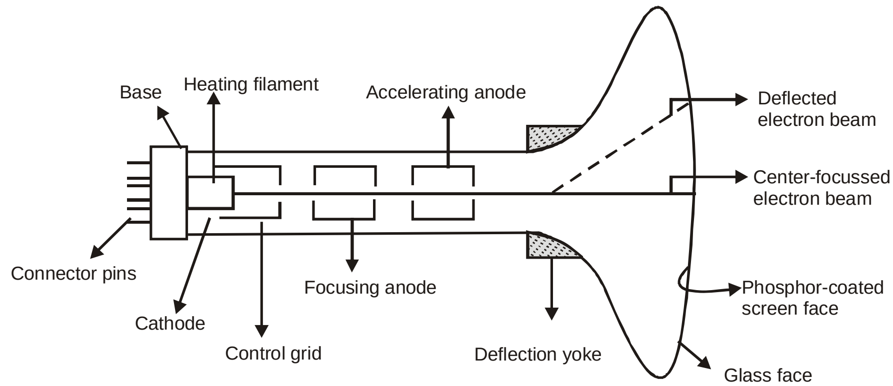


Fig. 1.2 *Schematic Diagram of a Raster Scan CRT*

When the electron beam converges on to a point on the phosphor-coated face of the CRT screen, the phosphor dots absorb some of the kinetic energy from the electrons. This causes the electrons in the phosphor atoms to jump to higher energy orbits. After a short time, these excited electrons drop back to their earlier stable state, releasing their extra energy as a small quantum of light energy. As long as these excited electrons return back to their stable state, phosphor continues to glow (phosphorescence) but gradually, it loses brightness. The time between the removal of excitation and the moment when phosphorescence has decayed to ten per cent of the initial brightness is termed as *persistence* of phosphor. The brightness of the light emitted by phosphor depends on the intensity with which the electron beam (number of electrons) strikes the phosphor. The intensity of the beam can be regulated by applying measured negative voltage at the control grid. Corresponding to a zero value in the frame buffer a high negative voltage is applied in the control grid, which in turn shuts off the electron beam by repelling the electrons and stopping them from coming out of the gun and hitting the screen. The corresponding points on the screen remain black. Similarly, a bright white spot can be created at a particular point by minimizing the negative voltage at the control grid of the three electron guns when the respective electron beams are directed to that point by the deflection mechanism.

Apart from the brightness, the size of the illuminated spot created on the screen varies directly with the intensity of the electron beam. As the intensity or number of electrons in the beam increases, the beam diameter and spot size increases. Also, the highly excited bright phosphor dots tends to spread the excitation to the neighbouring dots, thereby further increasing the spot size. Therefore, the total number of distinguishable spots (pixels) that can be created on the screen depends on the individual spot size. The lower the spot size, the higher the image resolution.

In a monochrome CRT, there is only one electron gun, whereas in a color CRT there are three electron guns each controlling the display of red, green and blue light respectively. Unlike the screen of a monochrome CRT, which has a uniform coating of phosphor, the color CRT has three color-phosphor dots (dot triad) – red, green and blue – at each point on the screen surface. When the red dot is struck by the electron beam, emits red light, the green dot emits green light and the blue dot emits blue light. Each triad is arranged in a triangular pattern, as are the three electron guns. The beam deflection arrangement allows all the three beams to be deflected at the same time to form a raster scan pattern. There are separate video streams for each RGB (Red, Green and Blue) color component which drive the electron guns to create different intensities of RGB colors at each point on the screen. To ensure that the electron beam

emitted from individual electron guns strikes only the correct phosphor dots (for example, the electron gun for red color excites only the red phosphor dot), a **shadow mask** is used just before the phosphor screen. The mask is a fine metal sheet with a regular array of holes punched in it. The mask is so aligned that as the set of three beams sweeps across the shadow mask they converge and intersect at the holes and then hits the correct phosphor dot; the beams are prevented or masked from intersecting other two dots of the triad. Thus, different intensities can be set for each dot in a triad and a small color spot is produced on the screen as a result. Figure 1.3 diagrammatically explains this concept.

NOTES

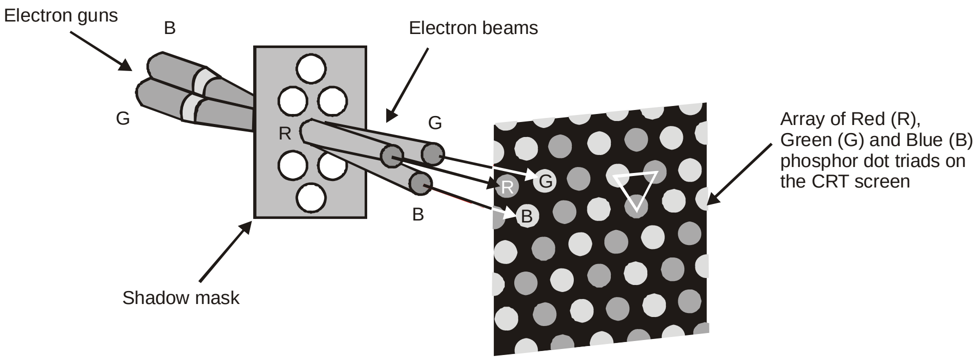


Fig. 1.3 *Electron Beams Passing through a Shadow Mask*

An alternative way to accomplish the masking function is also adopted by some CRTs. Instead of a shadow mask, they use an **aperture grill**. In this system, the metal mesh is replaced by hundreds of fine metal strips that run vertically from the top of the screen to the bottom. In these CRTs, the electron guns are placed side-by-side (not in a triangular fashion). The gaps between the metal wires allow the three electron beams to illuminate the adjacent columns of the colored phosphor which are arranged in alternating stripes of red, green and blue. This configuration allows the phosphor stripes to be placed closer together than conventional dot triads. The fine vertical wires block less of the electron beam than ordinary shadow masks, resulting in brighter and sharper image. This design is most common in Sony's popular *Trinitron*. Trinitron monitors are curved only in the horizontal plane but are flat vertically.

For TV sets and monitors the diagonal dimension is stated as the size. The edge of the picture tube is covered by the case, the actual viewable portion of the tube diagonally measuring only 19 inches. For standard monitors, the height is about three-fourth of the width. For a 19 inch monitor the image width will be 15 inches and the height will be 11inches.

Bit Planes, Color Depth and Color Palette

The appearance and color of a pixel of an image is a result of intersection of three primary colors (red, green and blue) at different intensities. When the intensities of all three electron beams are set to the highest level (causing each dot of a triad to glow with maximum intensity), the result is a white pixel; when all are set to zero, the pixel is black. Similarly, for many different combinations of intermediate intensity levels, several million color pixels are generated. For a mono monitor using a single electron gun, the phosphor material can glow with varied intensities depending on the intensity of the electron beam. As a result a pixel can be black (zero intensity) or white (maximum intensity) or have different shades of gray.

NOTES

The number of discrete intensities that the video card is capable of generating for each primary color determines the number of different colors that can be displayed. The number of memory bits required to store color information (intensity values for all three primary color components) about a pixel is called **color depth** or **bit depth**. A minimum of one memory bit (color depth = 1) is required to store intensity value, either 0 or 1 for every screen point or pixel. Corresponding to the intensity value 0 or 1, a pixel can be black or white respectively. So, if there are n pixels in an image, a total of n bits of memory used for storing intensity values will result in a pure black and white image. The block of memory which stores (or is mapped with) bilevel intensity values for each pixel of a full-screen pure black and white image is called a **bit plane or bitmap**. Figure 1.4 diagrammatically explain this concept when the bit depth = 1.

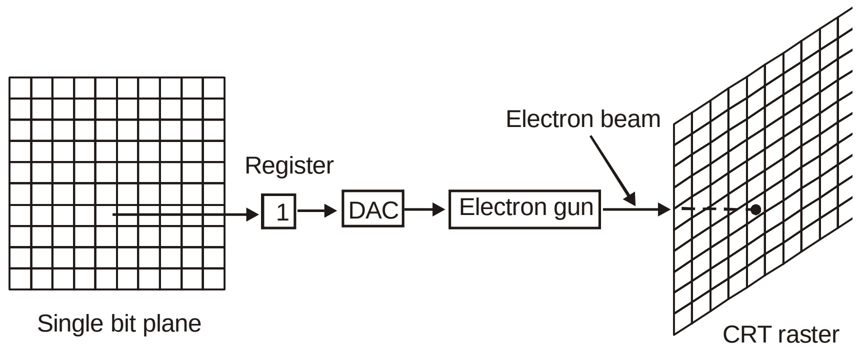


Fig. 1.4 *Pixel Illuminated by a Single (-) Bit Plane*

For bit depth =1, a pixel is illuminated (white) if intensity value 1 is stored in the corresponding memory address in the frame buffer.

Color or gray levels can be achieved in the display using additional bit planes. First consider a single bit plane – a planar array of bits, with one bit for each screen pixel. This plane is replicated as many times as there are bits per pixel, placing each bit plane behind its predecessor. Hence, the result for *n*-bits per pixel (color depth = *n*) is a collection of *n* bit planes that allows specifying any one of 2^n colors or gray shades at every pixel. This is explained in Figure 1.5.

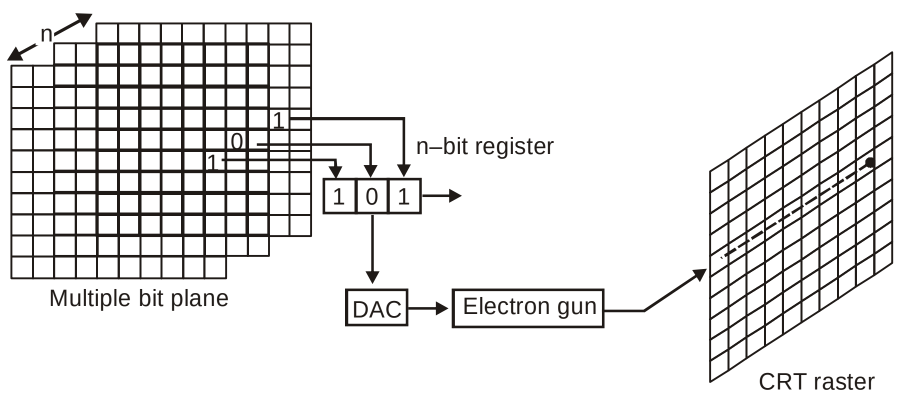


Fig. 1.5 *n Bit Planes Specify 2^n Colors*

For bit depth = *n* , *n* number of bit planes are used; each bit plane contribute to the gray shade of a pixel

The more the number of bits used per pixel, the finer the color detail of the image. However, increased color depths not only require significantly more memory

for storage, but also more data for the video card to process, which reduces the allowable refresh rate.

NOTES

Table 1.2 shows the general color depths.

Table 1.2 *Common Color Depths used in PCs*

Color Depth	Number of Displayed Colors	Bytes of Storage Per Pixel	Common Name for Color Depth
4-Bit	16	0.5	Standard VGA
8-Bit	256	1.0	256-Color Mode
16-Bit	65,536	2.0	High Color
24-Bit	16,777,216	3.0	True Color

For **True color** three bytes of information are used, one each for red, blue and green signals that comprise a pixel. A byte can hold upto 256 different values and so, 256 voltage settings are possible for each electron gun which means that each primary color can have 256 intensities, allowing more than 16 million (256 × 256 × 256) color possibilities. This facilitates in presenting a very realistic representation of the images, without necessitating any color compromise. In fact, 16 million colors is more than the human eye can discern. True color is a necessity for those doing high quality photo editing, graphical design, etc. Figure 1.6 shows the three 8-bit planes that are used to store colors of a pixel.

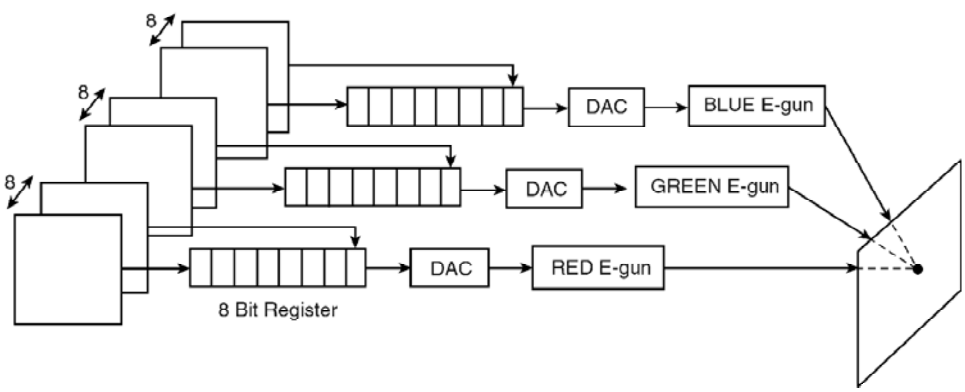


Fig. 1.6 *8 -Bit Planes for the Color of a Pixel*

For bit depth = 24 (true color display), 8 bit planes used for storing each primary color component of the color value of a pixel.

For **High color** two bytes of information are used to store the intensity values for all three colors. This is done by dividing 16 bits into three parts—5 bits for blue, 5 bits for red and 6 bits for green. This implies 32 (= 2^5) intensities for blue, 32 (= 2^5) for red and 64 (= 2^6) for green. This reduced color precision results in a loss of visible image quality, but one cannot easily see the difference between true color and high color image. However, high color is often used instead of true color because high color requires 33 per cent (or in some cases 50 per cent) in some cases less memory and also because image generation is faster.

In 256-color mode, the PC uses only 8 bits; this means something like 2 bits for blue and 3 each for green and red . There are chances that most of the colors of a given picture are not available and choosing between only 4 (= 2^2) or 8 (= 2^3) different values for each primary color would produce a rather blocky or grainy look of the displayed image. To resolve this problem a **palette or look-up table** is used.

A **palette** is a separate memory block (in addition to the 8 bit planes) created containing 256 different colors. The intensity values stored therein are not constrained within the range of 0 to 3 for blue and 0 to 7 each for green and red. Rather, each color is defined using the standard 3-byte color definition that is used in true color. Thus, the intensity values for each of the three primary color component can be anything between 0 and 255 in each of the table entries. Upon reading the bit planes, the resulting number instead of directly specifying the pixel color, is used as a pointer to the 3-byte color value entry in the look-up table. For example, if the color number, read from the bit-planes is 10 for a given pixel, then the intensities of red, green and blue to be displayed for that pixel will be found in the tenth entry of the table (see Figure 1.7). So the full range of true colors can be accessed, but only 256 of the available 16 million colors can be used at a time.

NOTES

Palette: A separate memory block (in addition to the 8 bit planes) created containing 256 different colors

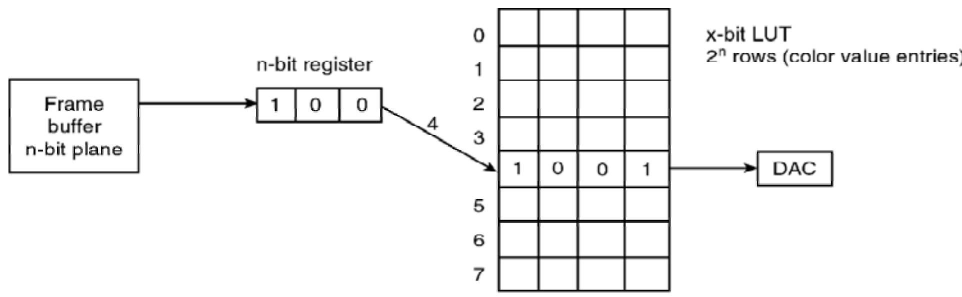


Fig. 1.7 *n-Bit Register Containing Pixel Intensity Value*

The *n*-bit register holds the row number of the look-up table; the particular row pointed contains the actual pixel intensity value which is a *x*-bit number ($x>n$)

The palette is an excellent compromise at the cost of moderate increase in memory: it allows only 8 bits of the frame buffer to be used to specify each color in an image and allows the creator of the image to decide what the 256 colors in the image should be. Because the palette can be reloaded any time with a different combination of 256 colors (out of 16 million) without changing the frame buffer values. Since virtually no image contain an even distribution of colors, this allows for more precision in an image by using more colors than would be possible by assigning each pixel a 2-bit value for blue and a 3-bit value each for green and red. For example, an image of the sky with clouds (like the Windows 95 standard background) would have different shades of blue, white and gray and virtually no red, green or yellow and their like.

256-color is the standard for much of the computing because the higher-precision color modes require more resource (especially video memory) and are not supported by many PCs. Despite the ability to 'hand pick' the 256 colors, this mode produces noticeably worse image quality than high color.

Frame Buffer and Output Circuitry

In the early days of PCs, the amount of information displayed was less. A screen of monochrome text, for example, needs only about 2 KB of memory space. Special parts of the Upper Memory Area (UMA) were dedicated to hold video data. As the need for video memory increased into the megabyte range, it made more sense to put the memory on the video card itself. In fact, to preserve the existing PC design limitations, it was necessary as there was simply no more space in the UMA to hold bigger screen images. The **frame buffer** is the video memory (RAM) that is used to hold or map the image displayed on the screen. The amount of memory required to hold the image

Frame buffer: The video memory (RAM) that is used to hold or map the image displayed on the screen

depends primarily on the resolution of the screen image and also the color depth used per pixel. The formula to calculate how much video memory is required at a given resolution and bit depth is quite simple:

NOTES

Memory in MB = (X resolution × Y resolution × Bits-per-pixel)/(8 × 1024 × 1024)

Practically, you need more memory than this formula can compute. One major reason is that video cards are available only in certain memory configurations (in terms of whole megabytes). For example, you cannot order a card with 1.7 MB of memory; you have to use a standard 2MB card available in the market. Another reason is that many video cards, especially high end accelerators and 3D cards, use memory for computation as well as for the frame buffer. Thus, they need much more memory than is required to hold the screen image.

Table 1.3 shows, in binary megabytes, the amount of memory required for the frame buffer for each common combination of screen resolution and color depth. The smallest industry standard video memory configuration required to support the combination is shown in the parentheses.

Table 1.3 *Video Memory Configurations*

Resolution	4 Bits	8 Bits	16 Bits	24 Bits	32 Bits
320 × 200	0.03 (256 KB)	0.06 (256 KB)	0.12 (256 KB)	0.18 (256 KB)	--
640 × 480	0.15 (256 KB)	0.29 (512 KB)	0.59 (1 MB)	0.88 (1 MB)	1.17 (2 MB)
800 × 600	--	0.46 (512 KB)	0.92 (1 MB)	1.37 (2 MB)	1.83 (2 MB)
1024 × 768	--	0.75 (1 MB)	1.50 (2 MB)	2.25 (4 MB)	3.00 (4 MB)
1280 × 1024	--	1.25 (2 MB)	2.50 (4 MB)	3.75 (4 MB)	5.00 (6 MB)
1600 × 1200	--	1.83 (2 MB)	3.66 (4 MB)	5.49 (6 MB)	7.32 (8 MB)

Some motherboard designs integrate the video chipset into the motherboard itself and use a part of the system RAM for the frame buffer. This is called **unified memory architecture**. This is done for cost saving. The result is almost much lower video performance, because in order to use higher resolutions and refresh rates, the video memory needs to have much higher performance than the RAM that is normally used for the system. This is also the reason why video card memory is so expensive as compared to regular system RAM.

In order to meet the increasing demand for faster and dedicated video memory at a comparable price, a technology introduced by Intel is fast becoming a new standard. It is called the **Accelerated Graphics Port** or (AGP). The AGP allows the video processor to access the system memory for graphics calculations, but keeps a dedicated video memory for the frame buffer. This is more efficient because the system memory can be shared dynamically between the system processor and the video processor, depending on the needs of the system. However, it should be remembered that AGP is considered to be a port – a dedicated interface between the video chipset and the system processor.

The display adapter circuitry on the video card or motherboard, in a raster graphics system, typically employs a special purpose processor called **Display Processor** or **Graphics Controller** or **Display Coprocessor** which is connected as an I/O peripheral to the CPU. Such processors assist the CPU in scan-converting the output primitives (line, circle, arc etc.) into bitmaps in frame buffer and also perform raster operations of moving, copying and modifying pixels or block of pixels. The output circuitry also includes another specialised hardware called *Video Controller* which actually drives the CRT and produces the display on the screen.

NOTES

The monitor is connected to the display adapter circuitry through a cable with 15-pin connectors. Inside the cable are three analog signals carrying brightness information, parallel for the three color components of each pixel. The cable also contains two digital signal lines for vertical and horizontal drive signals and three digital signal lines which carry specific information about the monitor to the display adapter.

The video controller in the output circuitry generates the horizontal and vertical drive signals so that the monitor can sweep its beam across the screen during raster scan. Memory reference addresses are generated in synchronization with the raster scan and the contents of the memory, are used to control the CRT beam intensity or color. Two registers (X register and Y register) are used to store the coordinates of the screen pixels. Assume that the y values of the adjacent scan lines increase by 1, in upward direction starting from 0 at the bottom of the screen to y_{max} at the top. Along each scan line the screen pixel positions, or *x* values are incremented by 1 from 0 at the leftmost position to x_{max} at the rightmost position. The origin is at the lower left corner of the screen, as is normal in a standard Cartesian coordinate system. At the start of a refresh cycle, the X register is set to 0 and the Y register is set to y_{max}. This (*x*, *y*) address is translated into a memory address of frame buffer where the color value for this pixel position is stored. The controller retrieves this color value (a binary number) from the frame buffer, breaks it up into three parts and sends each part to a separate Digital-To-Analog Converter (DAC). After conversion, the DAC puts the proportional analog voltage signals on the three analog output wires going to the monitor. These voltages in turn, control the intensity of the three electron beams that are focussed at the (x,y) screen position by the horizontal and vertical drive signals.

This process is repeated for each pixel along the top scan line, each time incrementing the X register by 1. As the pixels on the first scan line are generated, the X register is incremented through x_{max}. Then, the X register is reset to 0 and the Y register is decremented by 1 to access the next scan line. Pixels along this scan line are then processed and the procedure is repeated for each successive scan line until the pixels on the last scan line ($y = 0$) are generated. However, for a display system employing a color look-up table, the frame buffer value is not directly used to control the CRT beam intensity. It is used as an index to find the true pixel-color value from the look-up table. This look-up operation is done for each pixel on each display cycle. Figure 1.8 shows the general setup/architecture of a raster display system.

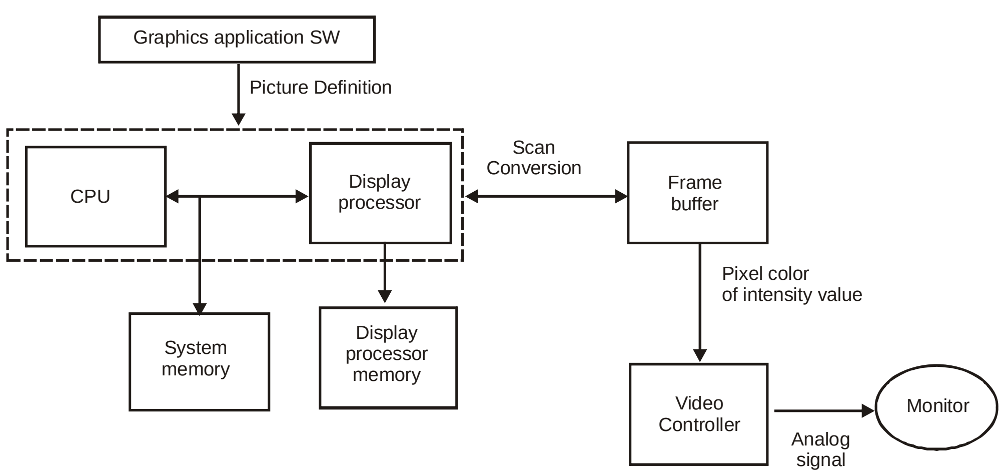


Fig. 1.8 *General Architecture of a Raster Display System*

NOTES

As the time available to display or refresh a single pixel in the screen is too less (few nanoseconds), accessing the frame buffer every time for reading each pixel intensity value would consume more time than what is allowed. Therefore, multiple adjacent pixel values are fetched to the frame buffer in a single access and stored in a register. After every allowable time gap (as dictated by the refresh rate and resolution), one pixel value is shifted out from the register to control the beam intensity for that pixel. This procedure is repeated with the next block of pixels and so on. Thus the whole group of pixels will be processed. This procedure has been illustration in Figure1.9.

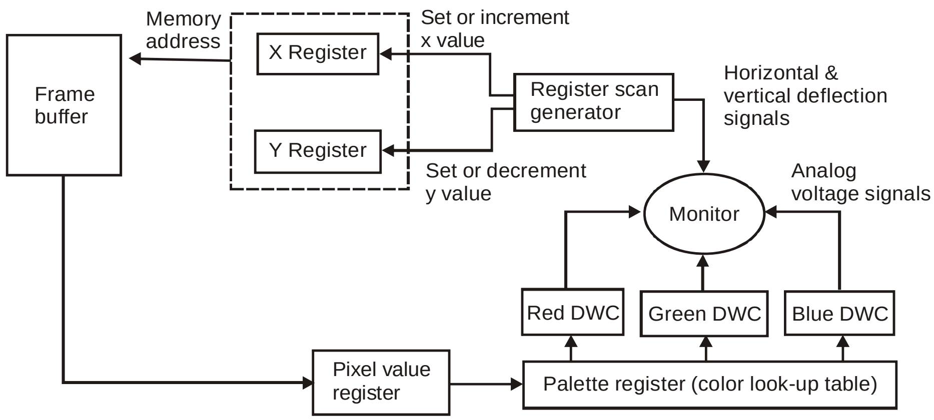


Fig. 1.9 *Logical Operations of the Video Controller*

Random Scan Display

There are basically two types of CRT's – Raster Scan type and Random Scan type. The main difference between the two is the technique with which the image is generated on the phosphor coated CRT screen. In raster scan method, the electron beam sweeps the entire screen in the same way as you would write a full page text in a notebook, word by word, character by character, from left to right and from top to bottom. Whereas, in random scan technique, the electron beam is directed straightway to the particular point(s) of the screen where the image is to be produced. It generates the image by drawing a set of random straight lines much in the same way one might move a pencil over a piece of paper to draw an image – drawing strokes from one point to another, one line at a time. This is why this technique is also referred as *vector drawing* or *stroke writing* or *calligraphic display*. Figure 1.10 (a-d) explain the random scan display technique.

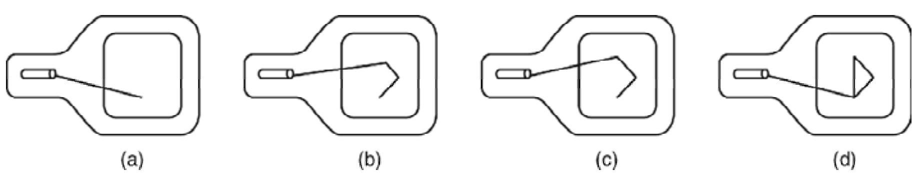


Fig. 1.10 *Drawing a Triangle on a Random Scan Display*

There are of course no bit planes containing the mapped pixel values in a vector system. Instead, the display buffer memory stores a set of line drawing commands along with the end point coordinates in a *display list* or *a display program* created by a graphics package. The Display Processing Unit (DPU) executes each command during every refresh cycle and feeds the *vector generator* with digital *x*, *y* and D*x*, D*y* values. The vector generator converts the digital signals into equivalent analog deflection voltages. This causes the electron beam to move to the start point or from

NOTES

the start point to the end point of a line or vector. Thus, the beam sweep does not follow any fixed pattern; the direction is arbitrary, as dictated by the display commands. When the beam focus must be moved from the end of one stroke to the beginning of the other, the beam intensity is set to 0.

Though the vector-drawn images lack in depth and real – like color precision, the random displays can work at higher resolutions than the raster displays. The images are sharp and have smooth edges unlike the jagged edges and lines on raster displays.

Direct View Storage Tube

Direct View Storage Tube (DVST) is rarely used today as a part of a display system. However, DVST marks a significant technological change in the usual refresh type display. Both in the raster scan and random scan system, the screen image is maintained (flicker free) by redrawing or refreshing the screen many times per second by cycling through the picture data stored in the refresh buffer. In DVST there is no refresh buffer; the images are created by drawing vectors or line segments with a relatively slow-moving electron beams. These beams are designed not to draw directly on phosphor but on a fine wire mesh (called *storage mesh*) coated with dielectric and mounted just behind the screen. A pattern of positive charge is deposited on the grid and this pattern is transferred to the phosphor coated screen by a continuous flood of electrons emanating from a separate flood gun (see Figure 1.11).

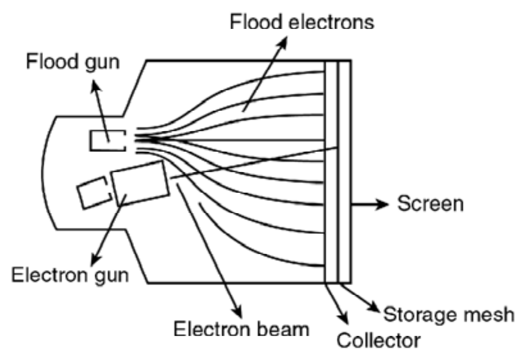


Fig. 1.11 *Schematic Diagram of a DVST*

Just behind the storage mesh, is a second grid, the *collector*, whose main purpose is to smooth out the flow of flood electrons. These electrons pass through the collector at a low velocity and are attracted to the positively charged portions of the storage mesh but repelled by the rest. The electrons that are not repelled by the storage mesh pass right through it and strike the phosphor.

To increase the energy of these slow moving electron and create a bright picture, the screen is maintained at a high positive potential.

The storage tube retains the image generated until it is erased. Thus, no refreshing is necessary, and the image is absolutely flicker free.

A major disadvantage of DVST is the interactive computer graphics inability to selectively erase parts of an image from the screen. To erase a line segment from the displayed image, one has to first erase the complete image and then redraw it by omitting that line segment. However, DVST supports very high resolution which is good for displaying complex images.

NOTES

Flat Panel Display

To satisfy the need of a compact portable monitor, modern technology has gifted us with LCD panel, Plasma display panel, LED panel and thin CRT. These display devices are smaller, lighter and specifically thinner than the conventional CRT and are thus termed as **Flat Panel Display (FPD)**. FPD in general, and LCD panels in particular, are most suitable for laptop (notebook) computers but are expensive to produce. Though the hardware prices are coming down sharply, the cost of the LCD or Plasma monitors are still too high to compete with CRT monitors in desktop applications. However, the thin CRT is comparatively economical. To produce a thin CRT the tube length of a normal CRT is reduced by bending it in the middle. The deflection apparatus is modified so that electron beams can be bend through 90 degrees to focus on the screen and at the same time can be steered up and down and across the screen. Figure 1.12 displays the interiors of a thin CRT monitor.

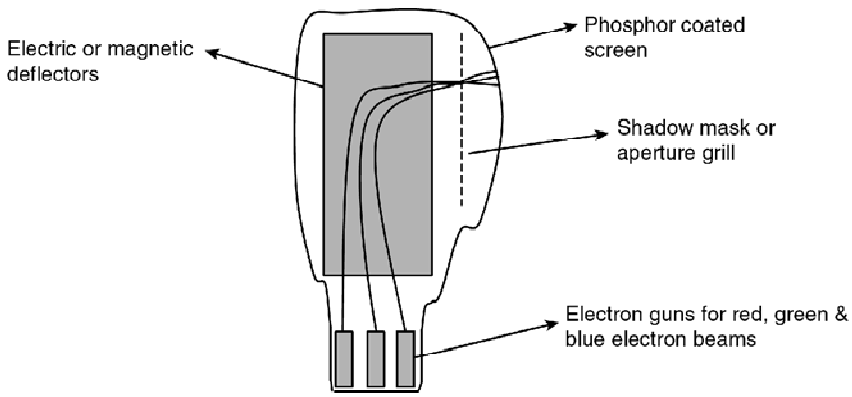


Fig. 1.12 *Thin CRT*

Liquid crystal Display or LCD

To understand the fundamental operation of a simple LCD, a model is shown in the Figure 1.13. LCD basically consists of a layer of liquid crystal, sandwiched between two polarizing plates. The polarizers are aligned perpendicular to each other (one vertical and the other horizontal), so that the light incident on the first polarizer will be blocked by the second. This is because a polarizer plate only passes photons (quanta of light) with their electric fields that are aligned parallel to the polarizing direction of that plate.

LCD: A flat panel display that uses the light modulating properties of liquid crystals

The **LCD** is a flat panel display that uses the light modulating properties of liquid crystals. displays are addressed in a matrix fashion. Rows of matrix are defined by a thin layer of horizontal transparent conductors, while columns are defined by another thin layer of vertical transparent conductors; the layers are placed between the LCD layer and the respective polarizer plate. The intersection of the two conductors defines a pixel position. This means that an individual LCD element is required for each display pixel, unlike a CRT which may have several dot triads for each pixel.

The liquid crystal material is made up of long rod-shaped crystalline molecules containing cyanobiphenyl units. The individual polar molecules in a nematic (spiral) LC layer are normally arranged in a spiral fashion such that the direction of polarization of polarized light passing through it is rotated by 90 degrees. Light from an internal source (backlight) enters the first polarizer (say horizontal) and is polarized accordingly (horizontally). As the light passes through the LC layer it is twisted 90 degrees (to align with the vertical) so that it is allowed to pass through the rear polarizer (vertical) and

then reflect from the reflector behind the rear polarizer. The reflected light when reach the viewers eye travelling in the reverse direction, the LCD appears bright (see Figure 1.13).

NOTES

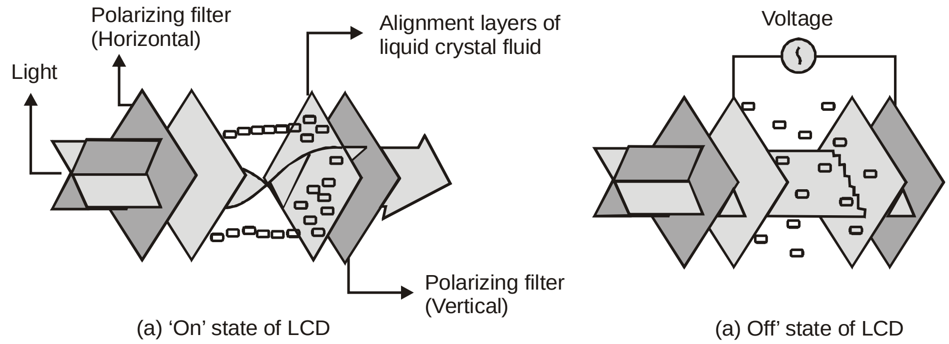


Fig. 1.13 *States of on LCD*

When an electric current passes through the LCD layer, the crystalline molecules align themselves parallel to the direction of light and thus, have no polarizing effect. The light entering through the front polarizer is not allowed to pass through the rear polarizer due to the mismatch of the polarization direction. The result is zero reflection of light and the LCD appears black.

In a color LCD, there are layers of three liquid crystal panels, one on top of another. Each one is filled with a colored (red, green or blue) liquid crystal. Each one has its own set of horizontal and vertical conductors. Each layer absorbs an adjustable portion of just one color of the light passing through it. This is similar to how color images are printed. The principal advantage of this design is that it helps create as many screen pixels as intersections, thus making higher-resolution LCD panels possible. In the true sense, each pixel comprises three color cells or sub-pixel elements.

The image painting operation in LCD panels is a different from that of the CRT though both are of raster scan type. In a simple LCD panel, an entire line of screen pixels is illuminated at one time. Then, the next line and so on, till the entire screen image is completed. Picture definitions are stored in a refresh buffer and the screen is refreshed typically at the rate of 60 frames per second. Once set, the screen pixels stay at fixed brightness until they are reset. The time required to set the brightness of a pixel is high as compared to that of the CRT. This is why LCD panel pixels cannot be turned on or off anywhere near the rate at which pixels are painted on a CRT screen. Except the high quality **Active Matrix LCD panels**, others have trouble displaying movies, which require quick refreshing.

Plasma Panel

Here a layer of gas (usually neon) is sandwiched between two glass plates. Thin vertical (column) strips of conductor run across one plate, while the horizontal (row) conductors run up and down the other plate. By applying high voltage to a pair of horizontal and vertical conductors, a small section of the gas (tiny neon bulb) at the intersection of the conductors breaks down into a glowing plasma of electrons and ions. Thus, in the array of gas bulbs, each one can be set to an 'on' (glowing) state or 'off' state by adjusting the voltages in the appropriate pair of conductors. Once set 'on', the bulbs remain in that state until explicitly made 'off' by momentarily reducing the voltage applied to the pair of conductors. Hence no refreshing is necessary.

Because of its excellent brightness, contrast and scalability to larger sizes, the plasma panel is attractive. Researches are on to eliminate the color-display limitation of such device at low production cost.

Input Devices

NOTES

Various devices are available for data input to general purpose computer systems with graphic capabilities or sophisticated workstations designed for graphics applications. Among these devices are graphic tablets, light pens, joysticks, touch panels, data gloves, image scanner, trackballs, digitizer, voice systems, the common alpha-numeric keyboard and mouse. The following sections discuss the basic functional characteristics and application of these devices.

1. Keyboard

Using a keyboard, a person can type a document, use keystroke shortcuts, access menus, play games and perform a variety of other tasks. Though keyboards can have different keys depending on the manufacturer, the operating system they are designed for and whether they are attached to a desktop computer or part of a laptop, still, most keyboards have between 80 and 110 keys, including the following:

- Typing keys (letters A to Z, a to z, characters like < , ? + = etc).
- A numeric keypad (numbers 0 to 9, characters like ! @ # ( ) etc).
- Function keys (F1 to F12).
- Control keys (Ctrl, Alt, Del, Pg Up, Pg Dn, Home, End, Esc, , Fn, arrow keys etc).

Function keys allow users to enter frequently used operations in a single keystroke. Control keys allow cursor and screen control. Displayed objects and menus can be selected using Control keys.

A keyboard is a lot like a miniature computer. It has its own processor, circuitry (key matrix) and a ROM storing the character map. It uses a variety of switch technology.

Though the basic working technology is same, there are design variations to make the keyboards easier and safer to use, versatile and elegant. Some of the non-traditional keyboards are Das keyboard, Virtual Laser keyboard, True-touch Roll-up keyboard, Ion Illuminated keyboard, Wireless keyboard (see Figure 1.14) etc.



Fig. 1.14 *Microsoft Wireless Keyboard*

2. Mouse

A **mouse** is basically a handheld pointing device, designed to sit under one hand of the user and to detect the movement that is relative to its two-dimensional supporting surface. It has become an inseparable part of a computer system just like the keyboard. There is a cursor in the shape of an arrow or cross-hair always associated with a

Mouse: A mouse is basically a handheld pointing device, designed to sit under one hand of the user and to detect the movement that is relative to its two-dimensional supporting surface

NOTES

mouse. You use the mouse whenever you want to move the cursor or activate something or drag and drop or resize some object on display. Drawing or designing figures and shapes using graphic application packages like AutoCAD, Photoshop, CorelDraw, Paint etc. is almost impossible without the mouse.

The mouse's 2D motion typically translates into the motion of a pointer on a display. In a **mechanical mouse** a ball–roller assembly is used; one roller used for detecting the X direction motion and the other for detecting the Y direction motion. An **optical mouse** uses LED and photodiodes (or optoelectronic sensors) to detect the movement of the underlying surface, rather than moving some of its parts as in a mechanical mouse. Modern **laser mouse** uses a small laser instead of a LED. Figure 1.15 shows a wireless mouse.



Fig. 1.15 *Microsoft's Two-button Wireless Mouse*

A mouse may have one, two or three buttons on the top. Usually clicking the primary or leftmost button will select items or pick screen-points and clicking the secondary or rightmost button will bring up a menu of alternative actions applicable to the selected item or specific to the context. Extra buttons or features are included in the mouse to add more control or dimensional input.

Trackball

A trackball is a pointing device consisting of a ball housed in a socket containing sensors to detect the rotation of the ball on two axes—like an upside-down mouse with an exposed protruding ball (see Figure 1.16). The user rolls the ball with their thumb, fingers or the palm of their hand to move the cursor. A potentiometer captures the trackball orientation which is calibrated with the translation of the cursor on screen. Tracker balls are common on CAD workstations for ease of use and, before the advent of the touchpad, on portable computers, where there may be no desk space on which to use a mouse.



Fig. 1.16 *A Logitech Trackball*

Joystick

A joystick is used as a personal computer peripheral or general control device consisting of a handheld stick that pivots about the base and steers the screen cursor around

(see Figure 1.17). Most joysticks are two-dimensional, having two axes of movement (similar to a mouse), but three-dimensional joysticks do exist. A joystick is generally configured so that by moving the stick to the left or right signals movement along the X-axis, and moving it forward (up) or back (down) signals movement along the Y-axis. In joysticks that are configured for three-dimensional movement, twisting the stick left (counter-clockwise) or right (clockwise) signals movement along the Z-axis. In conventional joysticks, potentiometers or variable resistors are used to dynamically detect the location of the stick and springs are there to return the stick to center position as it is released.



Fig. 1.17 *The Flighterstick, a Modern Programmable USB Joystick*

In many joysticks, optical sensors are used instead of analog potentiometer to read the stick movement digitally. One of the biggest additions to the world of joysticks is force feedback technology. On using a force feedback (also called haptic feedback) joystick, if you are shooting a machine gun in an action game, the stick would vibrate in your hands or, if you crashed your plane in a flight simulator, the stick would push back suddenly which means the stick moves in conjunction with, onscreen actions.

Joysticks are often used to control games and usually have one or more push-buttons whose state can also be read by the computer. Most I/O interface cards for PCs have a joystick (game control) port. Joysticks were popular throughout the mid-1990s for playing games and flight-simulators, although their use has declined with the promotion of the mouse and keyboard.

3. Digitizer and Graphics Tablet

A **digitizer** is locator device used for drawing, painting or interactively selecting coordinate positions on an object. Graphics tablet is one such digitizer that consists of a flat surface upon which the user may draw an image using an attached stylus, a pen-like drawing apparatus (see Figure 1.18). The image generally does not appear on the tablet itself, rather, it is displayed on the computer monitor.



Fig. 1.18 *A Tablet with Hand Cursor*

NOTES



Digitizer: Locator device used for drawing, painting or interactively selecting coordinate positions on an object

NOTES

The first graphics tablet resembling contemporary tablets was the RAND Tablet, also known as the Grafacon (for Graphic Converter), which an employed an orthogonal grid of wires under the surface of the pad. When pressure is applied to a point on the tablet using a stylus, the horizontal wire and vertical wire associated with the corresponding grid point meet each other, causing an electric current to flow into each of these wires. Since an electric current is only present in the two wires that meet, a unique coordinate for the stylus can be retrieved. The coordinate returned are tablet coordinates which are converted to user or screen coordinates by an imaging software. Even if it does not touch the tablet, the proximity of the stylus to the tablet surface can also be sensed by virtue of a weak magnetic field projected approximately one inch from the tablet surface. It is important to note that unlike the RAND Tablet, modern tablets do not require electronics in a stylus. Any tool that provides an accurate 'point', may be used with the pad. In some tablets multiple button hand-cursor is used instead of a stylus. Graphics tablets are available in various sizes and price ranges; A6-sized tablets being relatively inexpensive and A3-sized tablets being far more expensive.

Modern tablets usually connect to the computer via a USB interface. Because of their stylus-based interface and (in some cases) ability to detect pressure, tilt and other attributes of the stylus and its interaction with the tablet, they are widely used to create two-dimensional computer graphics. Free-hand sketches by an artist or drawing, following an existing image on the tablet are useful while digitizing old engineering drawing, electrical circuits, maps and toposheets for GIS. Indeed, many graphics packages, such as Corel Painter, Inkscape, Photoshop, Pixel image editor, Studio Artist, The GIMP are able to make use of the pressure (and, in some cases, stylus tilt) information generated by a tablet, by modifying attributes such as brush size, opacity and color. Three dimensional graphics can also be created by a 3D digitizer that uses sonic or electromagnetic transmissions to record positions on a real object as the stylus moves over its surface.

4. Touch Panel

A touch panel is a display device that accepts user input by means of a touch sensitive screen. The input is given by touching the displayed buttons or menus or icons with the finger. In a typical optical touch panel, LEDs are mounted in adjacent edges (one vertical and one horizontal). The opposite pair of adjacent edges contain light detectors. These detectors instantly identify which two orthogonal light beams emitted by the LEDs are blocked by a finger or other pointing device and thereby records the X and Y coordinates of the screen position touched for selection (see Figure 1.19). However, because of its poor resolution the touch panel cannot be used for selecting very small graphic objects or accurate screen positions.

The other two types of touch panels are electrical (or capacitive) and acoustical touch panel. In an electrical touch panel, two glass plates coated with appropriate conductive and resistive materials are placed face-to-face, similar to capacitor plates. Touching a point on the display panel generates force which changes the gap between the plates. This in turn, causes a change in capacitance across the plates that is converted to coordinate values of the selected screen position. In acoustic type, similar to the light rays, sonic beams are generated from the horizontal and vertical edges of the screen. The sonic beam is obstructed or reflected back by putting a finger in the designed location on the screen. From the time of travel of the beams, the location of the finger tip is determined.

Touch panels have gained wide acceptance in bank ATMs, video games and railway or tourist information systems.

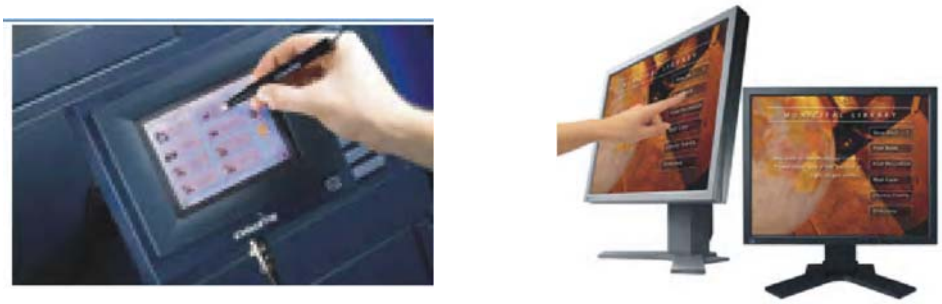


Fig. 1.19 Touch Panels

5. Light Pen

A **light pen** is a pointing device shaped like a pen and is connected to the computer. The tip of the light pen contains a light-sensitive element (photoelectric cell) which, when placed against the screen, detects the light from the screen enabling the computer to identify the location of the pen on the screen (see Figure 1.20). It allows the user to point to displayed objects or draw on the screen, in a similar way to a touch screen but with greater positional accuracy. A light pen can work with any CRT-based monitor, but not with LCD screens, projectors or other display devices.

The light pen actually works by sensing the sudden small change in brightness of a point on the screen when the electron gun refreshes that spot. By noting exactly where the scanning has reached at that moment, the X and Y position of the pen can be resolved. The pen position is updated on every refresh of the screen.



Fig. 1.20 Light Pen

Light pens are popularly used to digitize maps, create engineering drawing or storing signature or handwriting.

6. Data Glove

The data glove is an interface device that uses position tracking sensors and fiber optic strands running down each finger and connected to a compatible computer; the movement of the hand and fingers are displayed live on the computer monitor which in-turn allows the user to virtually touch an object displayed in the same monitor (see Figure 1.21). With the object animated it would appear that the user (wearing the

NOTES

Light pen: A pointing device shaped like a pen and is connected to the computer

NOTES

data glove) can pick up an object and do things with it just as he would do with a real object. In modern data glove devices, tactile sensors are used to provide the user with an additional feeling of touch or the amount of pressure or force the fingers or hands are exerting even though the user is not actually touching anything. Thus data, glove is an agent to transport the user to virtual reality.



Fig. 1.21 *Data Glove*

7. Voice System

The voice system or speech recognition system is a sophisticated input device that accepts voice or speech input from the user and transforms it to digital data that can be used to trigger graphic operations or enter data in specific fields. A dictionary is established for a particular operator (voice) by recording the frequency-patterns of the voice commands (words spoken) and corresponding functions to be performed (see Figure 1.22). Later, when a voice command is given by the same operator, the system searches for a frequency-pattern match in the dictionary and if found, the corresponding action is triggered. If a different operator is to use the system, then the dictionary has to be reestablished with the new operator's voice patterns.



Fig. 1.22 *Operator's Speech Recording*

8. Scanner

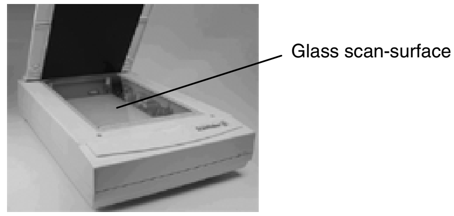
So far, some fundamental concept on how graphic images are generated and stored in some of the most common and widely used display systems have been discussed. Let us briefly study a graphic device which directly copies images from a paper or photograph and converts it into the digital format for display, storage and graphic manipulations. It is the scanner. Traditionally, the design and publishing houses have been the prime users of scanners, but the phenomenal growth of the Internet has made the scanner more popular even among the Web designers. Today, scanners are becoming affordable tools for the graphic artists and photographers.

NOTES

There are basically three types of scanners – **Drum, Flatbed** and **Sheetfed** scanners. Drum scanners are the high-end ones, whereas sheetfed scanners are the ordinary type. Flatbed scanners strike a balance between the two in quality as well as price. There are also *handheld scanners* or *bar-code readers* which are typically used to scan documents in strips of about 4 inches wide by holding the scanner in one hand and sliding it over the document.



(a) Scanner



(b) Scan-surface in a scanner

Fig. 1.23 *Flatbed Scanner*

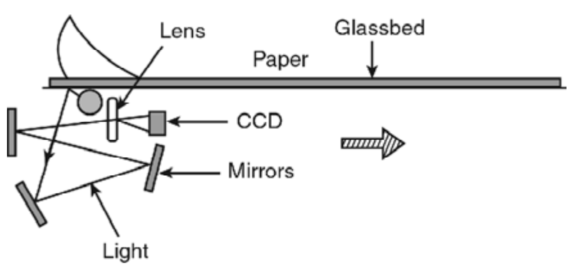
Flatbed Scanner

A flatbed scanner (Figure 1.23(a)) uses a light source, a lens, a Charge Coupled Device (CCD) array and one or more Analog-To-Digital Converters (ADCs) to collect optical information about the object to be scanned and transform it to an image file. A CCD is a miniature photometer that measures incident light and converts that into an analog voltage.

When you place an object on the copyboard or glass surface (Figure 1.23(b)) (like a copier machine) and start scanning, the light source illuminates a thin horizontal strip of the object called a raster line. Thus, when you scan an image, you scan one line at a time. During the exposure of each raster line, the scanner carriage (optical imaging elements, which is a network of lenses and mirrors) (Figure 1.24(a), (b)) is mechanically moved over a short distance using a motor. The light reflected is captured by the CCD array. Each CCD converts the light to an analog voltage and indicates the gray level for one pixel. The analog voltage is then converted into a digital value by an ADC using 8, 10 or 12 bits per color.



(a) Mirror and lens assembly in the scanner carriage



(b) Scanning operation

Fig. 1.24 *The Scanner Carriage and the Corresponding Scanning Operation*

The CCD elements are all in a row, with one element for each pixel in a line (see Figure 1.24). If you have 300 CCD elements per inch across the scanner, you can have a maximum potential optical resolution of 300 pixels per inch, also referred to as dots per inch (dpi).

NOTES

There are two methods by which the incident white light is sensed by the CCD. The first involves a rapidly rotating light filter that individually filters the red, green and blue components of the reflected light which are sensed by a single CCD device. Here, the color filter is fabricated into the chip directly. In the second method, a prismatic beam splitter first splits the reflected white light and three CCDs are used to sense the red, green and blue light beams.

Another imaging array technology that has become popular in inexpensive flatbed scanners is **Contact Image Sensor (CIS).** CIS replaces the CCD array, mirrors, filters, lamp and lens with rows of red, green and blue **Light Emitting Diodes** (LEDs). The image sensor mechanism, consisting of 300 to 600 sensors spanning the width of the scan area, is placed very close to the glass plate that the document rests upon. When the image is scanned, the LEDs combine to provide white light. The illuminated image is then captured by the row of sensors. CIS scanners are cheaper lighter and thinner, but do not provide the same level of quality and resolution that is found in most CCD scanners.

The output of a scanner is a bitmap image file, usually in a PCX or JPG format. If you scan a page of text, it may be saved as an image file which can not be edited in a word processing software.

Optical Character Recognition (OCR) softwares are intelligent programs which can convert a scanned page of text into editable text either into a plain text file, a Word document or even an Excel spreadsheet which can be easily edited. OCR can also be used to scan and recognize printed, typewritten or even handwritten text. The OCR software requires a raster image as an input, which may be an existing image file or an image transferred from a scanner. OCR analyses the image to find blocks of image information that resemble possible text fields and creates an index of such areas. The software examines these areas, compares shape of each object with a database of words categorized by different fonts or typefaces and recognizes the individual text characters from the information.

1.3.1 Multimedia Devices

Multimedia devices can be categorized into five classes, which are as follows:

1. **Capture Devices:** These devices are used to capture information. A video camera captures still as well as moving visual images in suitable formats. Scanners copy documents or images in image formats, such as *jpg, gif, bmp*, etc. Video recorder records captured images that are either still or in motion. Audio microphone captures sound. Other capture devices are keyboards, mouse, graphics tablets, 3D input devices, tactile sensors, Virtual Reality (VR) devices and digitizing/sampling hardwares.
2. **Storage Devices:** Such devices are used to store data as files of various formats. These devices are hard disks, optical storage devices, such as CD-ROMs, Jaz/Zip drives, DVD, pen drives, flash or thumb drives, etc.
3. **Communication Network Devices:** These devices establish communication between computers and also networks of computers and modems. These are Ethernet, token ring, FDDI, ATM, intranets and the Internet.

4. **Computer Systems:** Such devices are computer systems with peripherals supporting multimedia applications, networking and other devices. These are multimedia desktop machines, workstations, MPEG/VIDEO/DSP hardware, etc.
5. **Display Devices:** Such devices display outputs of different types. For sound, there are CD-quality speakers, for sound and video there are high-definition TV, SVGA monitors, hi-resolution monitors, color printers, etc.

   A typical multimedia system has been shown in Figure 1.25.



***Fig. 1.25** Multimedia System*

NOTES

Presentation Devices and the User Interface

Accelerated graphics ports, DVD, UDF file system are known as presentation devices that require configuration steps, hardware and software requirements with reference to multimedia technology. These presentation devices are configured with the operating system.

Now, the various presentation devices are used , which are already set with user interface. They are used as follows:

Accelerated Graphics Port (AGP)

AGP is considered to be a dedicated bus, which is used to deliver video and graphics output over Peripheral Component Interconnect (PCI) buses. Hence, it is considered to be helpful as a presentation device (see Figure 1.26).

AGP: Considered to be a dedicated bus, which is used to deliver video and graphics output over Peripheral Component Interconnect (PCI) buses

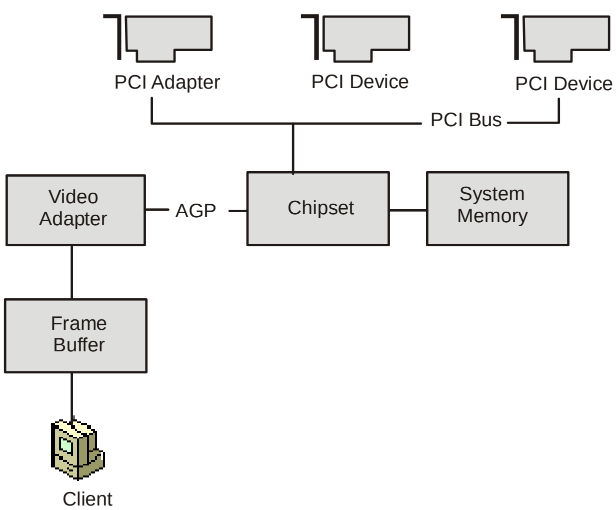


***Fig. 1.26** Accelerated Graphics Port Architecture*

NOTES

AGP comes with the following advantages over PCI video adapters:

- The bandwidth of AGP is four times more powerful than PCI. It has high sustaining rate because of splitting the transactions and sideband addressing.
- AGP, as a type of dedicated bus, lessens the contention of other devices.
- It makes the CPU to write into the shared system memory which is faster than directly writing in the local memory.
- The texture is read by AGP from the shared memory during the process of reading and writing data which is extracted from the local memory. That is why it improves the performance of high-resolution displaying technique of 3D graphics and scenes.
- AGP can run the graphics data directly from the system memory, instead of having to first move graphics data into video memory before running it.

However, for working with AGP video adapter, the system unit must be implemented with AGP graphics controller. A compatible chipset, such as the Pentium II LX or higher is also required to work with the AGP.

DVD

The DVD technology is able to work with data streams that are stored digitally. These data streams are concurrently used to play back multimedia applications as well as full-length motion pictures.

Capacity of DVDs: The current capacity of a DVD starts at 4.7 Gigabytes (GB). Both sides of the media can be readable and the data can be layered on each side (for example, a gold layer of data can be placed above a silver layer). Lower laser power is used to read the top layer and increased laser power allows the bottom layer to be read. Combining these two options increases the total possible capacity of a single DVD to 17 GB.

User Interface

User Interface (UI) is a system designed for the users. Developers develop the system to meet the user's needs. These systems are easy to use for the users.

UI can be defined as a complete combination of components and tools to make users interact directly with the computer system and computer control programs. The input devices, such as keyboard, mouse, touch pad, digital pen and output devices, such as display monitor, audio devices, touch screens, etc. are used by the users. User interface is basically a combination of graphical, audio and textual presentation of information by the program to a user and the users control sequence, i.e., punching of keystrokes, moving of the mouse and pressing the touch screen options by the user to control the program. UI is designed for various reasons such as the following:

- It is easy to use.
- It is less expensive to develop.
- It supports costs better user interface reduces the cost of training for the user.

- It maintains consistency in the user interface that enables it to lead the system as user-friendly.
- It maintains the easy rules that explain all the rules in details for each feature of the application step-by-step.
- It navigates the items which are defined in user interface.
- It words the messages and labels effectively.
- It helps in understanding the user interface widgets.
- It designs the color settings properly so that the used colors in the applications are consistently.
- It aligns the fields effectively.

User Interface Design Principle

The user interface design principle improves the quality of UI design. The principles are as follows:

- **Structure Principle:** This principle specifies the user's requirement. It puts related interface together, separates unrelated interface to make a user interface architecture in a friendly manner.
- **Simplicity Principle:** This principle follows simplicity to communicate. It also provides shortcuts to make the navigation easy.
- **Visibility Principle:** The options and materials are made visible without distracting the user's with redundant and extraneous information.
- **Feedback Principle:** This principle keeps the users well informed about actions, changes of state, errors, exceptions, etc. through concise, clear and relevant information.

Graphical User Interface (GUI)

GUI (pronounced as GOO-ee) is generally introduced in the user interface. The GUI includes the feel of multimedia parts, such as motion video, sound, virtual reality interfaces that become the part of GUI application used by the user. A system graphical user interface is sometimes referred to as 'look-and-feel'. The steps in user interface connect the command line interface. Sometimes non-graphical interface is used with GUI but it is generally not used by the users due to the inconvenient use of the keyboard commands, because users generally interact with the mouse instead of the keyboard commands. The non-graphical interface look good just because of the interfaces provided by the GUI. The Windows OS contains shell as a Graphical User Interface (GUI) in Microsoft Windows. The Windows shell contains windows components, for example, Start menu and taskbar. The first home GUI computer was declared as the Apple Mac released in 1984. Table 1.4 shows the basic GUI components.

NOTES

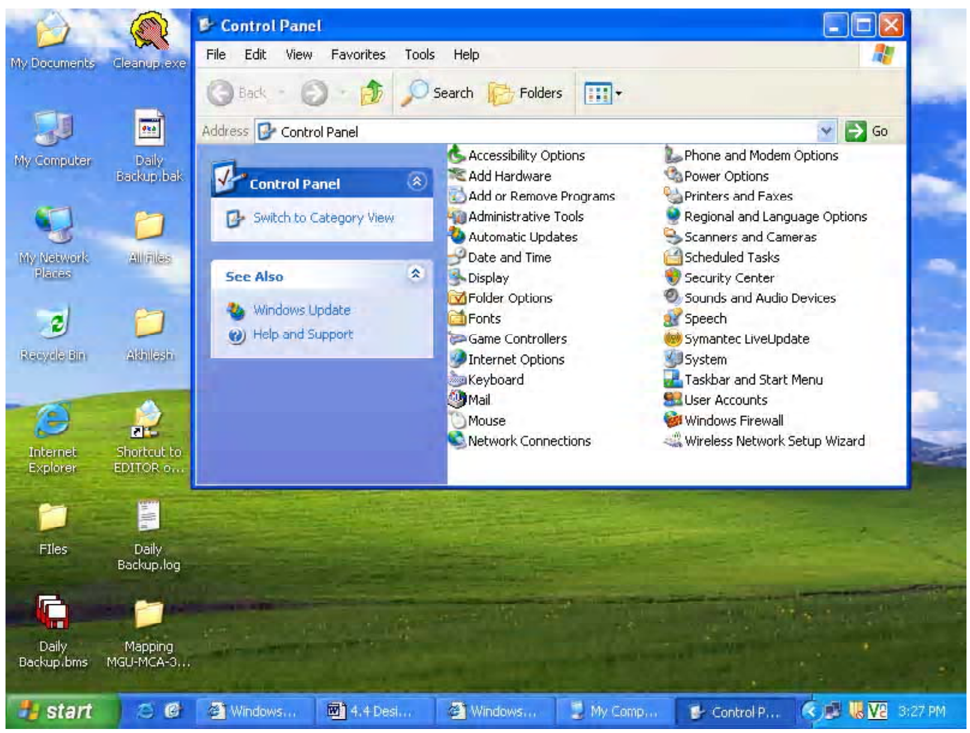
Simplicity principle: This principle follows simplicity to communicate. It also provides shortcuts to make the navigation easy

NOTES

Table 1.4 *Basic GUI Components*

Components	Function
Pointer	This tool is used to select the objects and commands.
Pointing device	Mouse, trackball etc. are used for indicating objects on the screen.
Icons	These tools are used as tiny pictures of commands, files, thumbnails, shortcuts, Windows, etc.
Desktop	This tool represents the area of display screen.
Windows	This prime tool represents all the information collectively in one area. Users can move the Windows around the display screen change the shape and navigate easily.
Menus	This tool gives a drop-down list of items so that users can easily switch with the operations.

The OS manages the resources in a system unit, such as disk drive, internal memory, mouse printers and network connections. The OS along with the GUI is a populated WIMP (Windows, Icons, Menus and Pointers) interface. The designers of OS keeps command shell for restricted uses, such as checking the system files and pinging the workstations to find network connectivity. A command shell is not a good user interface for frequent users as shown in screenshot below:



The following rules are followed to keep the user interface for multimedia presentation and applications:

1. Keep the user interface attractive and simple.
2. Maintain consistency.
3. Control the interaction (see Figure 1.27).

NOTES

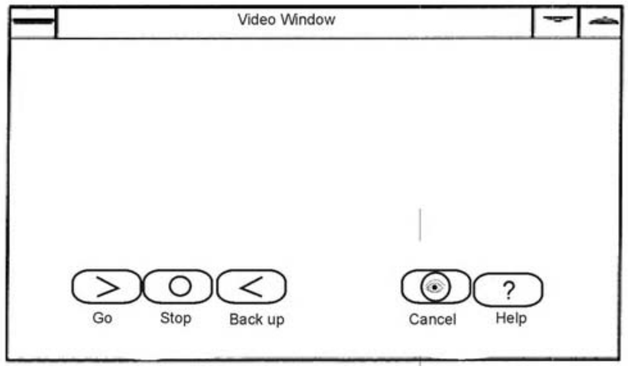


Fig. 1.27 *Button Available in Video Service*

4. Maintain the sound effect (see Figure 1.28).

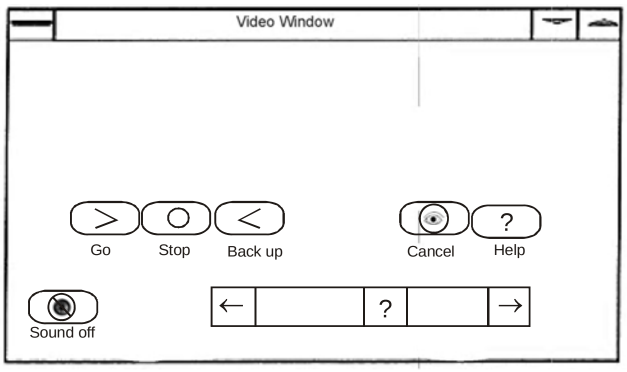


Fig. 1.28 *Sound Adjustment Option for Video Version*

5. Control the medium of touch (see Figure 1.29).

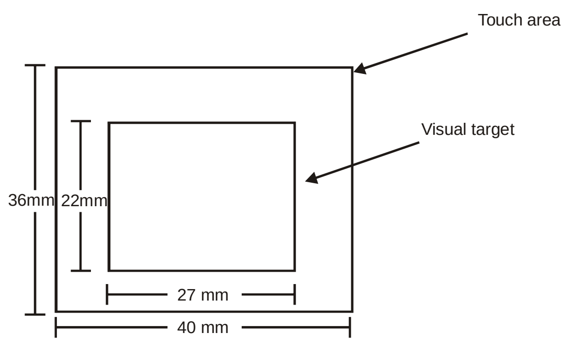


Fig. 1.29 *Area of Visual Targets*

1.3.2 Presentation Devices and the User Interface

The presentation devices are associated with the multimedia applications in GUI environment. GUI allows users to select the available resources visually and control the presentation devices, for example, a user can drag a slide thumbnail on top of any of the displays visible to show a slide on that display. It controls the virtual room lighting, hotspots, loudspeaker, printers, cameras, etc. The presentation device connects the computer using various wireless technologies, such as Bluetooth connection to produce the presentation online. The Bluetooth-enabled host computer is required for wireless connection. To produce the multimedia documents on the same network, a LAN wireless presentation device is used, whereas USB wireless presentation device

NOTES

is connected through the wireless networking service. For this, USB wireless presentation device is required to connect the wireless receiver into the available port that is assembled in the system unit. The speakers, shutter glasses, personal video recorder, headsets, speakers, high-definition multimedia interface, **accelerated graphics ports, DVD, UDF file system, etc.** are known as presentation devices that require the configuration steps, hardware and software requirements with reference to multimedia technology. The presentation devices are those devices that can be connected to the main components to the system unit, such as motherboard, RAM and ROM, CPU, etc. The other devices are also attached with mouse, keyboard, etc. The CPU status is maintained by control logic and offset address within 16 bits limit. The following working principles are stepped out by presentation devices:

- The DVD-ROM, CD-ROM, CD-R, USB flash device, sound cards, graphics cards, modem, docking station, network card, disk drives, scanners, microphones, etc. are the types of presentation devices that work with shared memory, i.e., global memory module.
- The devices use infra-red light sent from a variety of locations to control the signal.
- The presentation devices work individually, i.e., individual processor is connected with high-speed bus and connected to common input/output bus interface.
- Two different processing units complete the set of tasks with the help of various interrupt-driven devices, for example I/O devices and timers because they both determine the real-time working for CPU.
- These devices work on high input range, overload and the Internet protection, in case the system unit is connected with Internetwork connectivity.

These presentation devices are configured with the operating system. To configure the multimedia presentation devices, you will go through the following steps:

- Click on ***Sounds and Multimedia*** option in the Control Panel.
- Select ***Devices→Hardware*** option where you can select a device.
- The property of device is determined by Click ***Properties*** determing the property of the device, for example driver versions can be set using the Properties option.

The presentation devices are useful in streaming the multimedia files. Streaming audio video appeals users because they want to view it. In fact, streaming technology makes specialized videos and 3D animation in the field of entertainment, news, information, training, business, movies, TV show, etc. For example, in the news information technology, many users utilize the Internet to view video clips for domestic and international news items. The various presentation devices are explained as follows:

Shutter Glasses

Multimedia technology uses battery-powered active shutter glasses (see Figure 1.30), which delivers 3D images. However, these are very expensive as compared to polarized glasses. It includes 3D antiglare glasses. This battery powered glass is to cost about 12$ a pair, whereas the USB recharables costs 40$. Multimedia users need 3D TVs and 3D Blu-ray players. For example, the battery-powered glasses are positioned in a way so that ‘Monster vs Aliens’ presentation focuses 3D of set pieces for the first

time in multimedia. The 'Monster vs Aliens' is an American computer-animated 3-D feature film from Dreamworks animation and Paramount pictures.



Fig. 1.30 *Shutter Glasses*

NOTES

Personal Video Recorder (PVR)

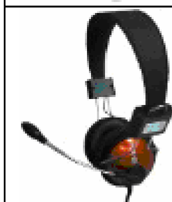
The PVR lets you record TV programming and also lets you pause and rewind the live broadcast videos and movies. It depends on the service you use and the access video-on-demand content, for example, SplitCam. This virtual video capture driver allows you to connect several applications to a single video capture source. It splits the video stream coming from the source and provides it to other client applications. You can connect upto 64 clients through a single video source. With this program, you get complete PVR functionality to pause live TV, rewind, fast-forward and conveniently record your shows for playback on your monitor or TV. It includes an integrated programming guide that provides a favorites manager, one-touch recording, search functionality and intelligent recording and scheduling. It also provides complete media center functionality to access your music, video and photo collections and playback DVDs, etc.

Headsets

Headsets: A prime presentable devices while using the multimedia files

The **headset** is considered as a prime presentable devices while using the multimedia files. The leather and metal finish puts engineered longevity and impression to make the desirable sound and view. For example, B&WP5 headphones are virtually leak-proof design that makes it a good choice for entering into the virtual world. Sound and voice, as it often the way, is slightly focussed into the more qualified success. Sometimes, marginal sound of iPods is quite obviously voiced to make the best choice. It also produces lovely, peppy and straight-edged low frequency music action, therefore, produced by the large scale dynamics and timing precision to come with outright communication. It even creates the pseudo-3D effect from 2-D content. Connecting laptop wirelessly brings audio quality extremely close from its USB socket. The traditional coaxial and optical devices support new Bluetooth protocol, which allows high-quality streaming with much better sound quality. The Bluetooth short range wireless link between two devices tends to be very robust, connection-free interference so that you can control the music by the headsets. The high-quality earphone with crystal clear sound and powerful bass is frequently used having high quality stereo earphone and powerful bass. The good headsets come with two extra ear buds having stereo earphone. The headphone impedance is about 32 Ohms, whereas output power is 3mW. Table 1.5 shows the various multimedia headsets and their functions:

Table 1.5 *Multimedia Headsets and their Function*

Types of Headsets	Function
	This headset is used for computer/mp3/mp4 and Hi-Fi stereo sound.
	This headset is having steel framework and microphone. It is used for network chatting and telephonic conversation.
	This headset is soft, comfortable, efficient and convenient and is used for stereo headphone to enjoy great music across the net.
	This headset is attractive, durable and user-friendly. It is used for chatting with other users.
	This headset is used in laboratory language system to monitor the networking performance in the transmission of multimedia files and documents.
	This Hi-Fi music headset is used in the wide frequency range and provides low distortion level. It is in fact, widely used for MP3 and MP4 player.

Speakers

The speakers are used to entertain the music or to enable hearing the sound watch the DVD. The speakers (see Figure 1.31) are used to hear music and watch the DVD and also listen to music or watch DVD. The multimedia enjoyment devices are portable in nature and extra space is required for that. The speakers, known as circular device, are used in multimedia devices that facilitates the 360 degree and hence can be projected the sound in certain direction. The digital player can be plugged or hooked with mega speaker phone. This is supported by Bluetooth compatibility.



Fig. 1.31 *Speaker*

NOTES

High-Definition Multimedia Interface (HDMI)

The HDMI device (see Figure 1.32) contains high-quality digital audio and video connection. For this, Ethernet cables are underpinned to connect the network-enabled bit, such as computers, printers, routers, game consoles and Audio Visual (AV) receivers, etc. For example, the HDMI1.3 handles HD audio formats. HDMI incorporates with digital right management and provides an interface between compatible digital audio/ video sources, such as system unit, set top box, video monitor, digital television, video game system, DVD player, etc. HDMI device is a small, user-friendly interconnect that can carry up to 5 Gbps of combined video and audio in a single cable. This system eliminates the cost, complexity and confusion of multiple cables used to connect current AV systems. This interface transmits pure digital video and digital audio over a single, easily managed cable. Like the DVI standard on which it is based, HDMI transmits uncompressed high-definition video with a theoretical bandwidth over 5 Gb/sec, but adds 6-channel digital audio and a bi-directional control channel that allows the components to communicate with each other.



Fig. 1.32 *HDMI*

High-Capacity Hard Disk Device

An external hard disk drive (see Figure 1.33) is required for computer system if the hard disk is insufficient to store all the important multimedia data and documents. First, you need to learn the quick installation of this peripheral. For configuration, you need to connect the mini-connector of the USB cable to the back of the drive. Then, connect the standard USB port to your computer. The line labelled with symbol provides auxiliary power. It is needed only if your computer does not provide the sufficient USB power through the other line. For PC users, the drive icon should appear in 'My Computer' or 'Window's Explorer'. But, for Mac user, you would need to reformat the drive. The external and portable hard disk is used to support the back up storage. It takes the safe periodic system backups frequent data backups and also protects the critical data. The frequency of backup depends on the importance of data. To protect the data, you need to simply drag and drop copies of the critical files onto the drive. Any data storage device can fail so you must always keep at least two copies on different disks for all types of multimedia files and documents. The two types of messages, such as *'Drive not detected'* and *'PC will not start up'* are appear on the screen if the hard disk drive is configured incorrectly with the PC during the time of troubleshooting process.

NOTES



Fig. 1.33 *High-Capacity Hard Disk Device*

Digital Versatile Disc (DVD)

DVD is also known as an important presentation device and that is able to work with data streams that are digitally stored. These data streams are concurrently used to playback multimedia applications as well as full-length motion pictures. For example, portable region-free DVD player is frequently used for multimedia applications. It is of a 15" widescreen and multi-format digital media playback. This is a full-featured portable multimedia player with a large Liquid Crystal Display (LCD) screen providing a crisp display alongwith a feature set that effortlessly combines music, video and image viewing into a stylish unit that can be used at home, in the office, etc.

MP3 Players

The MP3 players these days come with satellite radio capability that can store up to 50 hours of content (1 GB) storage or even up to 25 hours (512M**B**). It is sometimes supported by radio home docking kit that allows the device to receive XM Satellite Radio at home. It is possible to store digital content with this player. You can see video and photo viewing features on what used to be a simple portable music player. The stored content can be played anywhere for a totally portable listening experience (see Figure 1.34).



Fig. 1.34 *MP3 Player*

This convergence seems likely to continue, which is why it supports both, music and multimedia. This player allows users to create and manage customized playlists, combining both personal digital music files and recorded XM programming. The other accessories used in this player are battery, earbuds, home dock/cradle, home XM antenna, A/C power adapter, RCA cable, USB cable, etc. MP3 decoding is easier than encoding, but requires decompression and filtering. The basic CD is used standard for data discs. Figure 1.35 illustrates the CD/MP3 Player arrangement.

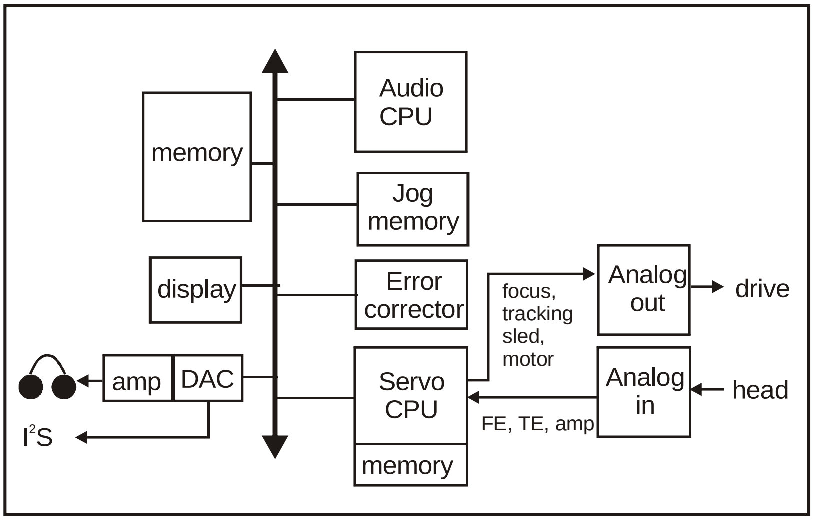


Fig. 1.35 *CD/MP3 Player Arrangement*

MP3 is used on data DVD (stereo) and also for audio track of video DVD. Dolby digital (AC3) audio stream was developed by Dolby Labs that allows five separate audio channels. These channels hold a subwoofer channel along with left and right rear, left and right front and centre, for example Dolby Digital 5.1-channel audio is a discrete multi-channel surround sound system. Discrete means that the sound information contained in each of the six available channels is distinct and independent from the others.

Configuration of Sound Schemes and Sound Events

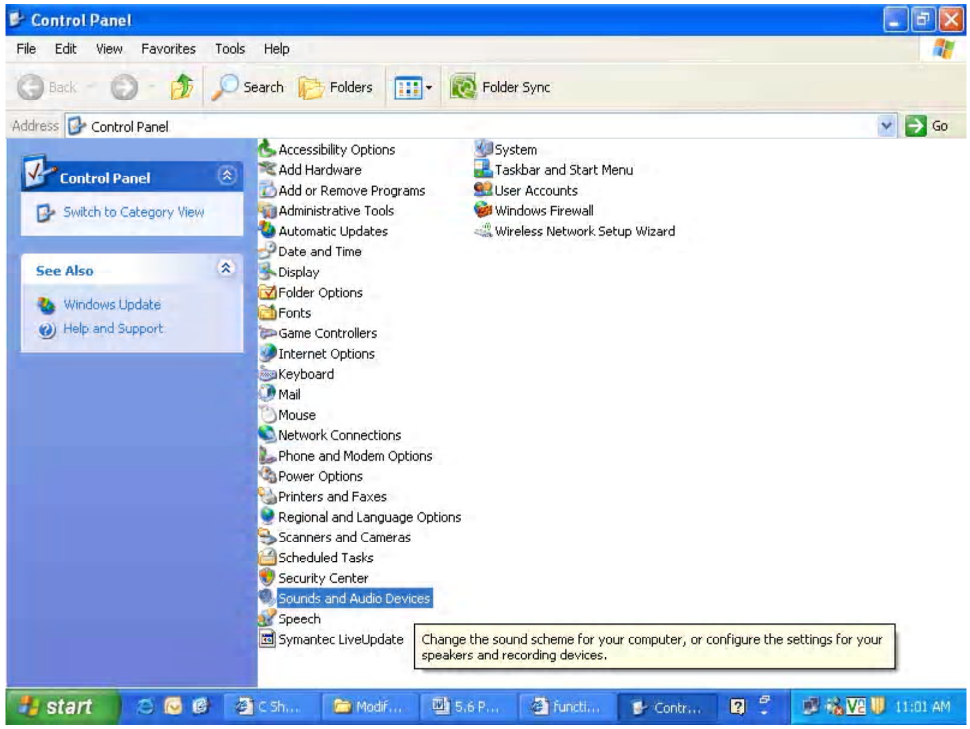
Use the following procedures to configure event/sound pairings and save them as custom sound schemes:

To Assign Sounds to System and Program Events

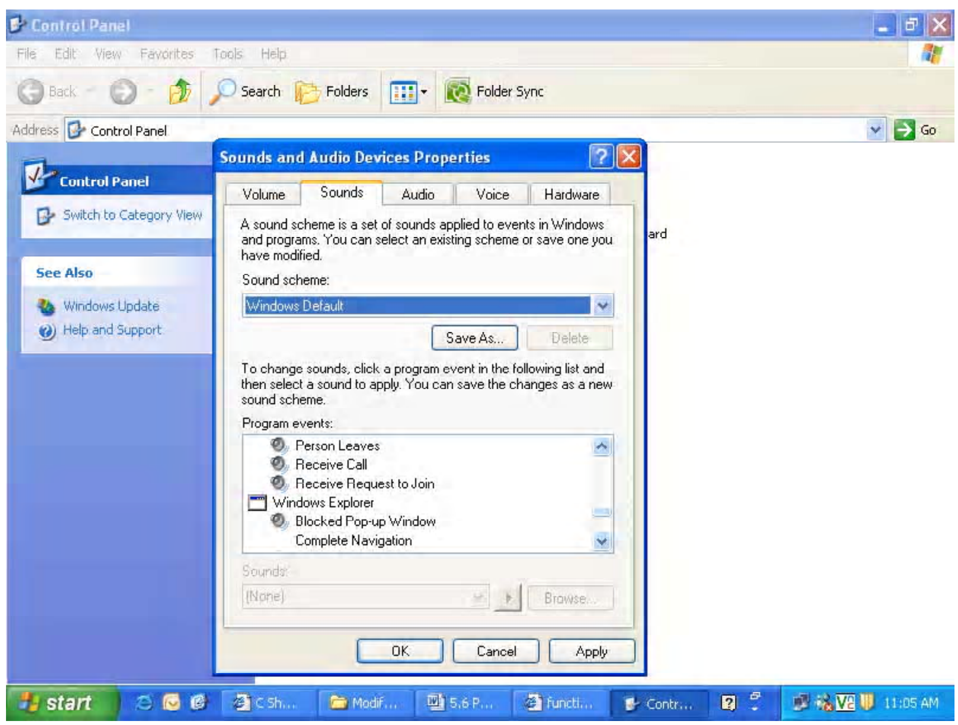
Perform the following steps to install sound system and program events to the system unit:

The first step is to open the Control Panel and select '**Sounds and Audio Devices'** option in the Windows XP platform. This is shown in the following screenshot.

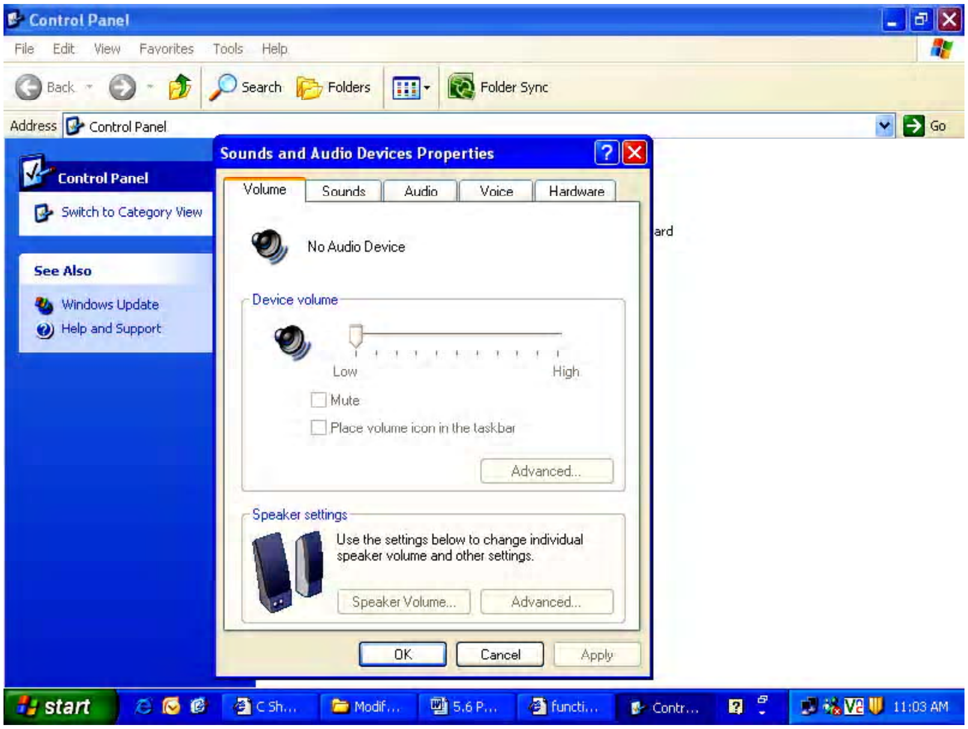
NOTES



The Sounds and Audio Devices Properties tab appears as shown in the following screenshot.



Click the **Sounds** tab to set the system or program event to which you want to assign a sound. The preceding screenshot shows the sounds tab.



NOTES

If you want to play the sound files, which are not listed, then select the ***Browse*** option that can be accessed to the sound files on your computer or network. You can also select a different sound scheme in the ***Scheme*** box. You can save an entire set of sound and event pairings as a custom sound scheme.

Configuration of Recording and Playback Devices

The PC can have multiple audio recording and playback devices. These preferred devices are used to record or play the sound. The recording volume can also be set in the PC. The following steps are followed:

To Configure a Preferred Sound Playback or Recording Device (Windows 2000)

1. Open the Control Panel to select **Sounds and Multimedia** option.
2. **Audio** tab provides **Sound Recording or Playback** option. The preferred device is selected to record or play a sound.
3. Select the **Volume** to set the volume controls.
4. Select **Advanced** → **Performance** tab to set the properties after achieving the hardware acceleration level.

To Configure Audio Performance Options

The audio performance can be configured by stepping out of the following steps:

1. Select **Control Panel** → **Sounds and Multimedia**.
2. Select **Audio** tab. In the **Sound Recording or Sound Playback** option, select **Advanced** button**.**
3. In the **Performance** tab select **Audio Playback** or **Audio Recording.** Here you can set quality of sample rate conversion and hardware acceleration for the system unit.

The IE plays videos, animations and sound by default from the Internet and intranet sites. However, you can disable any of these options to ensure that the pages load faster or to enhance a quiet work environment. You can also configure Internet Explorer to play a specific radio station by default, every time the browser starts.

To Enable or Disable Sounds, Videos, and Animation From Web Pages:

1. Select **IE** → **Properties**.
2. Select **Advanced** tab. Click on **Multimedia** section to use the check boxes for setting **Play Sounds, Play Animations** or **Play Videos**.

Wireless Devices and Multimedia

The various wireless devices and technologies, such as Bluetooth and Wi-Fi connections are required to correlate with presentation devices and user interfaces. They are used for multimedia documents and files in the following ways:

Bluetooth

Bluetooth is used in wireless Personal Area Networks (PANs). It connects and exchanges the information between devices, such as mobile phones, laptops, personal computers, printers, digital cameras and video game consoles via a secure, globally unlicensed short-range radio frequency. The word 'Bluetooth' is derived from the 10th century Danish King Harald Bluetooth. The Bluetooth technology has been designed to connect both mobile devices and presentations that currently require a wire. The services of Bluetooth are provided by wireless in which no setup is needed. The transfer speed of data rate is 1.0 Mbps. It holds radio frequency chips, which are plugged into the devices. It maintains 2.45 GHz frequency. It uses a technique called spread-spectrum frequency hopping. A wireless solution is given for the Bluetooth to reduce use of cables net in the presentations. It is a type of replacement for Infrared Data Association (IrDA). It is also considered as complementary technology for Apple Airport and valid for 802.11b, 802.11g and 802.11n. It is about 2.4 GHz radio technologies which are used to eliminate the cables between various devices. These devices are referred to as computers, laptops, phones, mouse, printers and other equipments. The bi-directional radio transmission is used to deliver the automatic wireless connections. Bluetooth is a standard for tiny, radio frequency chips that are plugged into the devices. These chips are designed to take all of the information that wires are normally sent and transmitted at the special frequency by Bluetooth chip.

Wi-Fi

Wi-Fi stands for Wireless Fidelity. It is used for wireless devices. Wi-Fi Multimedia (WMM) formerly known as 'wireless multimedia extensions refer to QoS (Quality of Service) over Wi-Fi. For home entertainment and especially for Network Attach Storage (NAS) boxes, the Wi-Fi connection is involved. QoS enables Wi-Fi access points to prioritize traffic and optimizes the way shared network resources are allocated among different applications. The Wi-Fi alliance represents the wireless standard protocol and basically non-profit organizations. This supports interoperability features for wireless devices. It connects the networking system without cables. But for this, Wi-Fi and regular ISP services are needed. The manufacturers of Wi-Fi alliance build various devices for 802.11 standards. Approximately 205 companies joined to the Wi-Fi alliance and almost 900 products have been certified to the interoperable system.

These companies give assurance that the Wi-Fi devices are connected by physical layer in reference models. Wi-Fi Protected Access Solution (WPA) was recently added to the Wi-Fi standard. The physical and access control layer implement the extra enhanced features, such as Internet security. The Wi-Fi can be grow in leaps and bounds because it is connected via a spectrum. It uses unlicensed 24 GHz and 5 GHz bands. It provides data throughput for most uses. The prime equipment is required for Wi-Fi connection is Wi-Fi PC card. It is common way to connect to the computer to the Internet without wires. This card is technically known as Personal Computer Memory Card International Association (PCMCIA). A hotspot is used in Wi-Fi connection to mean an area in which Wi-Fi users connect to the Internet. The Wi-Fi hotspots are created around the antennas to outlet the radio waves of wireless networking. It is confined to almost 10,000 hotspots in crowded areas, such as airport lounges, cafes, etc. A series of antennas are set up into the city-wide zones. The Internet connection is facilitated by Wi-Fi chips. The long calls possible in Wi-Fi are by bypassed network and VoIP technologies. The Wi-Fi assembled mobiles and laptops are connected to these hotspots frequently. The amount for these is paid after using this technology by credit card on the login page provided by the Web browser. Users can hold their accounts provided by service providers, such as BT Openzone, Skypezone, Ninetendo Wi-Fi, T-Mobile, O2, etc. The Wi-Fi access point is interconnected with Wi-Fi devices that are configured with router. Modem is also used to make connection between router and Internet DSL cable. The main role of router is to connect the Wi-Fi access point and wired network clients. The Voice-over IP (VoIP) software enables data, fax calls and voice across IP networks and represents Internet telephony allowing a communication between two PCs over packet switching Internet. It works by encoding voice information. Then, it is changed to digital format. It provides cost benefits by converging data and voice over IP network into the mobile phones. Many of the latest mobiles are connected with Wi-Fi via VoIP technology. Between the Internet connection and Wi-Fi access point there needs to be hardware designed to connect with the Internet and share the internetworking connectivity. Figure 1.36 shows how a Wi-Fi Zone is made up.

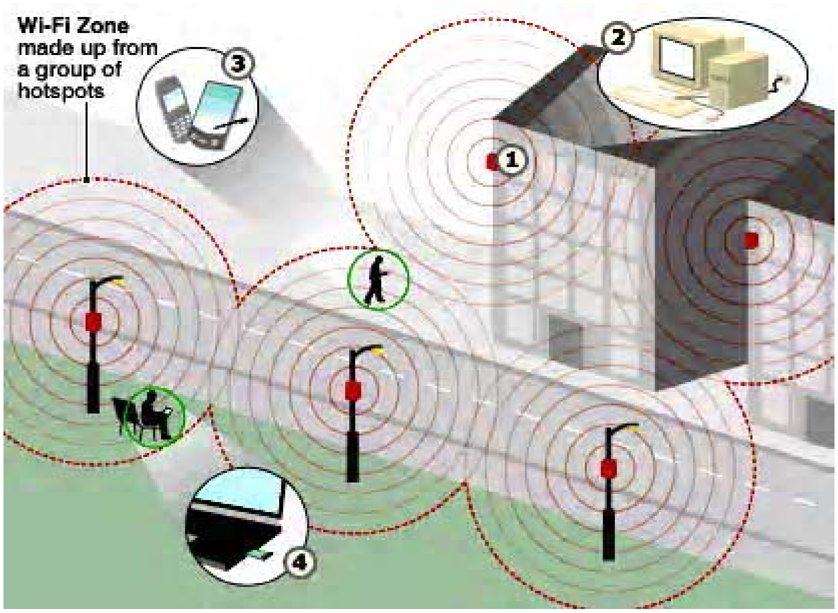


Fig. 1.36 *Wi-Fi Zone*

NOTES

User interface: Designed for the system to the users

User Interface

User interface is designed for the system to the users. The developers develop the system to meet the user's need are easy to use for users. UI is designed in terms of people, users that need to interact with the system for a specific device, machine and computer program. The two prime relevant tools are used in the following way:

- Input, allows users to input the data and information through keyboard or by pointing devices,
- Output, is known as displayed tool and allows the system to produce the result as per the inputted information or data.

Common User Interface (CUI)

Multimedia user interfaces combine several kinds of media to help people use a computer, which is termed as a common user interface. These media can include text, graphics, animation, images, voice, music and touch. Multimedia user interfaces are gaining popularity because they are very effective at keeping the interest of their users, improve the amount of information users remember and are very cost-effective. It keeps the user interface simple. A CUI is a combination of Graphical User Interface (GUI) and Command Line Interface (CLI). This interface combines the functionality of the GUI as well as the flexibility and power of a CLI in the same application. Both CLIs and GUIs have advantages. CLIs provide speed and power to adept users. GUIs offer a less complicated, more intuitive approach to computer interaction. As GUI-based operating systems, such as Windows and the applications provided on these platforms have developed and have incorporated more and more functionality. Generally, as a GUI begins to offer more functionality and options, it becomes slower and less intuitive to use as the user is offered too many choices. CLIs in operating systems such as MS-DOS and UNIX contain a huge number of textual commands, arguments and attributes (appendages or options for the command entered) in their command language. These commands needed to be entered in an exact prescribed form in order to achieve the desired result, resulting in many errors and frustration for users. While knowledge of every single command and its correct use is not necessary (and most likely impossible) for operation of a CLI, a certain level of expertise is needed for competent usage.

- **Keep the User Interface Attractive and Simple:** This rule guides you to show off the multimedia technology but in practical it works.
- **Maintain Consistency:** Users must use similar objects that performs better services foe the multimedia technology.
- **Control the Interaction:** Make the user interface with the help of presentation devices. For example, while watching a video, you can cancel the video or user can select the footage to watch or replay the video, etc.
- **Sound Effect:** User can make the sound effect softer or louder for personal convenience. User can select push button at this stage and change button to inverse the video. The system unit can give a short beep that is to be processed after pushing the buttons.

- **Medium of Touch:** The touch medium is helpful for users to be prompted by choices, such as product navigation decisions can have various choices and can be selected by issuing the commands. The 40 mm wide and 36 mm high is suggested for touch area that gives the visual targets for push buttons. The push buttons can be showed on the window as 27 mm wide and 22 mm high.

A graphical user interface for use in connection with computer display systems, such as computer controlled multi-media editing systems. The interface utilizes the components of color, for example hue, luminance and saturation to convey information to a user. Each of these components are mapped to variables that are displayed via the interface. The value of a particular variable may be represented by a gradient of one of the color components or by a discrete value of one of the color components.

1.4 MULTIMEDIA PLATFORMS

Platforms come in many shapes and sizes. When telecommunications access and core networks are concerned, they must be defined clearly in terms of content as well as characteristics. This means that the choices of technology and configurations must be verified in a controlled environment. A true multimedia platform integrates and combines various multimedia devices and components.

Multimedia applications concept spans a wider range of services by which one cannot adequately define a multimedia platform for all applications.

Factors that may influence your choice of multimedia development platform include:

- **Platform-dependence** – Both hardware and OS.
- **Programming language** – C/C++, Java, .NET languages.
- **Functionality** – Graphics, streaming media.
- **Deployment** – PC/PDA, local or networked, Web deployment.

Some of the known multimedia platforms are Apple Macintosh and IBM Compatible PC.

Many of the multimedia applications that we run on our PC, to play games, watch video, or browse through the photos, require some version of Direct X to be installed. DirectX 9 SDK includes:

- **Direct3D** – for graphics, both 2D and 3D.
- **DirectInput** – Supporting a variety of input devices.
- **DirectPlay** – For multiplayer networked games.
- **DirectSound** – For high performance audio.
- **DirectMusic** – To manipulate (non-linear) musical tracks.
- **DirectShow** – For capture and playback of multimedia (video) streams.

In addition to above multimedia components for the platform, there is an API for setting up these components.

NOTES

Check Your Progress

1. Define the term multimedia.
2. How is image quality related with resolution?
3. What is the use of capture devices?
4. What is the function of personal video recorder in a presentation device?
5. What factors are involved in the digital representation of sound?
6. What is the use of a headset in the multimedia environment?
7. What are multimedia authoring tools used for?
8. List the various types of multimedia software.
9. Give examples of multimedia platforms.
10. What do you mean by authoring?
11. How hypertext differs from hyperlink?
12. What do you mean by cross platform?
13. What do you mean by recorded multimedia presentation?

1.5 DEVELOPMENT TOOLS: TYPES

NOTES

Authoring System:
A collection of software tools that help in various aspects of multimedia production.

Authoring Tools

The key to successful multimedia production is a seamless integration of multimedia elements for graphic design, content management, production and packaging. The whole process of developing a multimedia package is called authoring. An **authoring system** is a collection of software tools that help in various aspects of multimedia production. Multimedia elements are woven together using authoring tools. These tools are designed to manage individual multimedia elements and provide user interaction. Multimedia authoring software can be used to make:

- Video productions.
- Animations.
- Games.
- Presentations.
- Interactive kiosk applications.
- Interactive training.
- Simulations, prototypes and technical visualisations.

Steps in Multimedia Production

The following steps can be followed in multimedia production:

1. ***Media Capture:*** Multimedia authoring systems streamline data capture by providing interfaces to a range of image and data capture devices.
2. ***Media Conversion:*** Images, audio clips, animation, sequences and video clips exist in a variety of formats. A well-equipped multimedia authoring system will include a set of utilities for converting between many of the commonly used formats.
3. ***Media Editing:*** After data has been captured and converted to the native format of the authoring system, it may need some polishing before it is suitable for presentation. For instance, 'noise' can be removed from audio clips, images can be touched up, etc. Multimedia authoring systems provide media – specific editors for these operations.
4. ***Media Composition:*** The core of a multimedia authoring system includes a tool for combining media and specifying their spatial (one image being juxtaposed or placed side-by-side within a second) and temporal (when an audio track is added to a visual sequence) relationships.

1.5.1 Elements of Multimedia

The following are the various elements of multimedia:

1. Text

Text, containing words and symbols, is the most common form of communication. It is one of the popularly used mediums of appearance that is used to deliver information accurately and in detail. Usually text provides the core structure to the package. Words are vital elements of multimedia that can appear in the titles, menus, navigation aids

and in the content of a multimedia application or project. It is most essential to use words that have the most precise and powerful meanings to express what you need to convey.

A major drawback of using text is that it is not user friendly as compared to the other elements of multimedia. For example: it is harder to read from a screen for long, as it tires the eyes more than reading it in its print version.

NOTES

2. Designing with Text

From a design perspective, the choice of font size, style and other text attributes needs to be related both to the complexity of the message and to its venue.

Some useful tips for designing the text in your multimedia application:

- Use legible fonts that can be easily read.
- Vary the font size and style according to the importance of your message.

AAAAA

- Indent your paragraphs wherever required.
- Explore the effects of different colors and shadows to add depth to your application.
- Use menus for easy navigation and meaningful words for menu items.
- Use buttons, icons or symbols for user interaction.

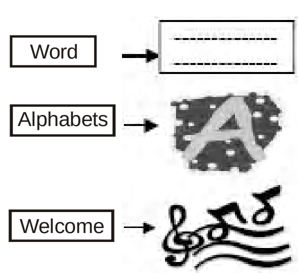


- You can also use stylish fonts for displaying attention-grabbing results.



3. Hyperlinks

Hypertext is the organisation of information units into connected associations that a user can choose to make. An instance of such an association is called a **Hyperlink.** When a user clicks on such a link, more information on the particular topic is displayed. It, therefore, provides the user an option of reading as much information as required. Hyperlinks can contain cross-linking of words not only to words but also to images, videos or sound files. Hyperlinks are used for non-linear navigation, which is not an option available in a sequentially organized book.



NOTES

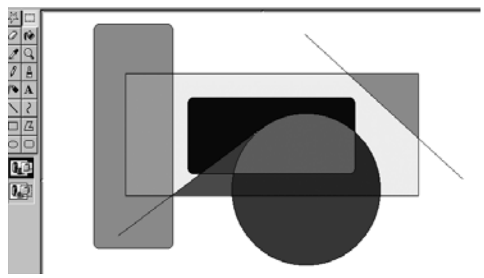
4. Graphics

Pictures/graphics enhance the overall look of a multimedia package. Pictures express more than normal text and are generally considered as the most important element of a multimedia application.

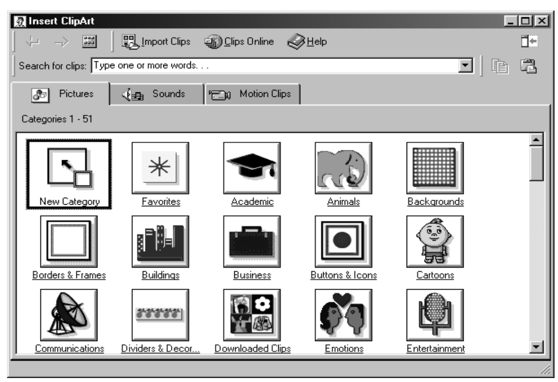
It is often noticed that a Webpage containing numerous images takes longer to download than a simple text based webpage. Image files are, therefore, compressed to save memory and disk space of your computer. **GIF** (Graphics Interchange Format), **JPEG** (Joint Photographic Experts Group) and **PNG** (Portable Network Graphics) are examples of compressed image file formats.

Pictures can be created using any of the following ways:

- You can use drawing tools like MS paint to create simple pictures. Paint allows to create or assemble pictures by drawing straight, wavy or curved lines, using shapes like squares, circles and polygons or simply by freehand drawing.



- You can insert images from the ClipArt gallery. A ClipArt collection typically contains a series of images for different categories. ClipArt is available through CD-ROMs or from the Internet.



- You can use scanner (Figure 1.37(b))or digital camera (Figure 1.37(a)) to capture original pictures in digitised form. You can also scan images, created using traditional methods like watercolours, crayons etc.



(a) Digital Camera **(b) Scanner**

Fig. 1.37 *Capturing of Original Pictures in Digital Form*

- If you are still not satisfied with scanned pictures or images downloaded from the ClipArt gallery, you can use image editing tools to manipulate images according to your taste. Picture properties that can be manipulated using these tools include brightness, contrast, color, depth, hue and size. Apart from correcting the blemishes in your image, you can add additional effects like filters, shadows, patterns, 3D-effects and many more. Image editing tools are indispensable for excellent multimedia production. Adobe Photoshop, Microsoft PhotoDraw are some examples of image editing tools.

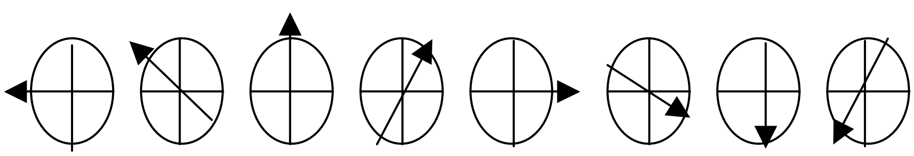
Some useful tips for adding images in your multimedia application:

- Do not overload your application with too many images. This would not only make your application look clumsy, but it would also consume excessive computer resources.
- Do not include heavy (oversized) graphics.
- Use context – sensitive images.
- To the extent possible, use compressed image file formats. Avoid using unoptimised graphics.
- Pick the right color or combination of colors for your multimedia application. Color scheme used for the text should blend well with the images.

1.5.2 Animation

Animation gives visual impact to your multimedia application. In simple terms, it can be defined as an entity moving across the screen. This entity could be a text object or an image. An animation consists of a series of rapidly changing objects, which when blended together gives an illusion of movement. The speed with which each object is replaced by the next one is so rapid that the eye perceives this as motion.

Example: Consider the process of rotating a wheel, the position of the arrow changes so rapidly that it gives an illusion of spinning.



Animation Tools: MS PowerPoint is a tool used for creating primitive animations. Visual effects like wipes, dissolves, fades and zooms can be added to any object. For example: you can make a text to fly from top or left. Such effects are available with almost all authoring packages. You can create complex animations using tools like Director, 3D Studio Max, CompuServe and Shockwave. Such animations can be ported across platforms and applications by making use of suitable translators.

Some useful tips for adding animations in your multimedia application:

- Before you create an animation, organise its execution into a series of logical steps. First choose the objects in your presentation that you want to animate and then decide the sequence of animation. In case of complicated presentations, writing a detailed script of the list of activities will prove useful.

NOTES



Animation: It gives visual impact ot your multimedia application. In simple terms, it can be defined as an entity moving across the screen

NOTES

- You can animate one or all objects of an application. As mentioned earlier, applications intended for the Web should not contain too many animations as it would affect the download time.
- Add user interactivity wherever essential to the application.
- You can combine animations with live sounds for catching the user's attention.

1.5.3 Sound

Sound is used to set the rhythm or a mood in a package. Speech gives an effect of a language (pronunciation) for instance. Proper usage of sound can make all the difference between an ordinary multimedia presentation and a professional one.

Sound Types: Musical Instrument Digital Interface (**MIDI**) is a communication standard developed in early 1980's for electronic musical instruments and computers. A MIDI file consists of a list of commands that represent the recordings of musical actions. When these commands are sent to a MIDI Playback device, a sound is produced. The main disadvantage of MIDI data is that it is not digitised. In contrast, **Digital Audio** is a recording that depends on the capabilities of your sound system. Digital audio data are the actual representations of sound data and are stored in the form of thousands of individual numbers known as **Samples**. Digital sound is used for music CDs.

Creating Sound: Sound can be recorded using a microphone, a synthesiser, or any other medium like tape or cassette player and then be digitised using audio digitising software. Therefore, sound may be digitised from any source – natural or pre-recorded. Digitised sounds are stored as wave (.WAV) files (Windows Platform). These can then be played using Windows Media player.

Some useful tips for adding sounds to your multimedia application:

- A distorted recording sounds terrible, so before a sound file is added it must be tested for clarity. If required, it must be edited using Audio-Editing tools like Wave Studio Sound or Macromedia's Sound Edit 16-2. It may be worth noting that higher the sound quality, the larger would be the file size.
- Decide the kind of sound you need, such as background music, special sound effects or spoken dialogue.
- Test the sound, to ensure they are synchronised properly with images and or animations.

1.5.4 Video

If pictures can paint a thousand words, then motion pictures can paint a million. Digital video is the most engaging of multimedia venues and is a powerful tool for bringing users close to the real world.

NTSE (National Television Standards Committee), PAL (Phase Alteration Line), SECAM (Sequential Color and Memory), HDTV (High Definition Television) are commonly used broadcast and video standards across the globe.

Current television is based on analog technology and fixed international standards for the broadcast and display of images. Computer video, on the other hand is based on digital technology and other standards for displaying images. Digital Video is

produced using analog video as a base. The conversion of analog video into its digital equivalent requires a special hardware called **Video Capture Card.**

Video data is also compressed using different compression techniques. MPEG (Motion Pictures Expert Group), JPEG (Joint Photographic Experts Group), P*64, real video are examples of commonly used compressed video formats.

NOTES

Some useful tips for adding video clips to your multimedia application:

- Video clippings which are not appropriately designed can degrade your presentation rather than add value to it. Carefully planned, well-executed clips can make a dramatic difference in a multimedia presentation.
- Titles used in video clips should be plain enough to be easily read.
- Avoid making busy title screens; use more screen if required.
- Again, any multimedia element that is added to an application intended for Web should be compressed to support quick and easy download.

1.5.5 Cross Platform Compatibility

In computer technology, cross platform or multi-platform refers to the unique characteristic of computer software which enables it in implementing methodologies for inter-operating on several computer platforms. Typically, the cross platform software can be classified into following two types:

- First that supports specified creation or compilation for each platform that it holds.
- Second that can be directly run on any platform without any specification, such as software written in an interpreted language and precompiled portable bytecode which are universal and standard components supported by all platforms for the interpreters or run-time packages.

The cross platform compatibility specifies the software or hardware capability that can run indistinguishable on different platforms. For example, nowadays numerous applications for the Windows and the Macintosh have the ability to create binary compatible files which indicate that users can switch from one platform to the other without changing the data into a new format because the data will be supported due to cross compatibility feature. Hence, the cross platform computing is gaining significant importance since local area networks use advanced technology and are compatible to link computer systems of different types.

In multimedia, the cross platform is the unique capability of an application that helps in accurately performing on a range of computers, operating systems and Web browsers. Basically, it includes the following:

- Hardware Issues.
- Operating Systems.
- Environment.
- Software.
- Elements.
- Design.
- Publication.

NOTES

Hardware Issues: It includes the following:

- **Device Type:** PC/Laptop (including Tablet PCs)/Mobile Devices (Phones, PDAs).
- **Device Performance:** Based on the performance of Processor/Hard Drive/Optical Drive.
- **Available Peripherals:** Mouse, Keyboard, Stylus (PDAs), Screen (Size, Touch Screen), Speakers, Microphone (for Voice Input), Hand Controllers, Joysticks.
- **Media:** CDs/DVDs, Internet Connection (Broadband/Dial-Up)
- **Monitor/Screen Resolution for a PC:** 2005 - 800 × 600 Pixels, 2010 – 1024 ´ 768 Pixels, 2o11 - Superior Web Solutions.

Nowadays maximum numbers of people use their mobile devices to surf the Internet which are equipped with advanced technologies and software.

Operating Systems: It includes the Windows/Macintosh/UNIX. Some of the ‘old’ multimedia applications are specifically designed for different screen resolutions and quality of color. Hence, must be run in compatibility mode or on a virtual machine.

Environment: It includes the following:

- Internet versus CD-ROM/DVD
  - o **Graphics:** File Sizes, Types, such as gif, jpg, png on Web.
  - o **Video:** File Sizes, Codec Availability.
  - o **Updates:** For CD-ROM - Fixed, For Internet - Easily Changed.
  - o **Errors:** For CD: To Remaster, For Internet - Easily Changed.
- To build an iPhone application requires an Apple Computer, Intel only.
- Web Issues: Corporate firewall policies may restrict access to certain sites, such as game sites and social networking sites which are often blocked.

Software: It includes the following:

- **Browser Issues:** Browser Specific Tags (I.E.- Marquee), HTML-5 Multimedia Tags and JavaScript capabilities, such as Google Chrome Experiments.
- **Plug-Ins:** Adobe Flash Player Version, Microsoft Silverlight
- **Application Specific:** Flash capable to create standalone exe/swf files or html/swf files.
- **Supported File Types:** The following file types are supported in multimedia on the Internet:
  - o **Images:** On Web - gif, jpg, png and animated png.
  - o **Videos:** Microsoft WMV Format for Windows, DivX (usually wav), Apple Quicktime (mov) and MPEG.

Elements: It includes the following for the Web:

- **Image Size:** Use thumbnails for images and also uses file compression (jpg), and resize the image.
- **Music:** Background sound may clash with foreground sounds or user preferred music.

- **Music Formats:** It includes MIDI, WMA (not Linux), MP3 (Linux).
- **Font Choices:** Different fonts for PC/Mac/Linux, UNIX, for example, Arial, Helvetica, Times New Roman, Calibri, etc.

Design: It includes the following:

- Different conventions for different systems.
- Different interfaces, for example Google Android and Microsoft Windows Phone 7.
- Colors can be handled differently in different systems.
- Web page colors may vary in different Web browsers on different platforms. The 216 of 250 possible index colors are safe for use.
- Different available default fonts. Text can be converted to a graphic or CSS font families that are currently used.
- Guides to develop an application to the lowest common specification.
- Different languages. Text is easier to convert though may require a language pack to support special characters, such as Chinese, Japanese, Russian, etc.

Publication

- CD-ROM/DVD requires burning to CD/DVD, producing labels and distributing which includes marketing.
- Internet requires uploading to Web server usually using FTP.
- Internet based applications require a technology infrastructure, as follows:
  - o Choice of Web Host.
  - o Cost of Uploading/Downloading Files and Capacity.

HTML Multimedia

Typically, multimedia on the Web is pictures, sound, music, films, videos and animations. Nowadays, the Web browsers support various multimedia formats because the recent Web pages have embedded multimedia elements.

Browser Support: The first Internet browsers only supported text which was limited to a single font in a single color. Later the new browsers were developed which had support for different colors, fonts and text styles and also support for pictures. Various browsers support sounds, animations and videos in different ways. Some directly support multimedia elements whereas some need an extra helper program, such as a plug-in.

Multimedia Formats: Multimedia elements, such as sounds or videos are stored in media files. The most universal way to determine the type of a file is to check the file extension. When a browser observes the file extension as .htm or .html, it considers the file as an HTML file. The .xml extension specifies that it is an XML file and the .css extension specifies that it is a style sheet file. Pictures are distinguished by extensions .gif, .png and .jpg. Multimedia files have their specific formats with extensions .swf, .wav, .mp3 and .mp4.

Multimedia formats: Multimedia elements, such as sounds or videos are stored in media files

NOTES

1.5.6 Commercial Tools

Multimedia presentation is of two types, live or recorded. In recorded multimedia presentation, the interaction is possible through a navigation system while in a live multimedia presentation the interaction is done with the help of a presenter or performer. The example of a live multimedia performance is a laser show. There are various commercial tools of multimedia technological or digital multimedia technology that enhance the users' experience to convey information, for example technology involving illusions of taste and smell. The following are some examples of commercial tools that can be used to develop a multimedia presentation:

Drawing, Painting and Graphic Tools

Adobe Photoshop: The Photo Editing Tool

Adobe Photoshop is a graphics editing program developed and published by Adobe Systems. The new advanced version of Adobe 2003 'Creative Suite' is reframed, rebranded and named as Adobe Photoshop CS. Adobe Photoshop CS6 is the 13th major release of Adobe Photoshop. Adobe Photoshop is released in two specific editions: Adobe Photoshop and Adobe Photoshop Extended with the Extended extra 3D image creation, motion graphics editing and advanced image analysis features. In 2011, a version of Adobe Photoshop was specifically released for the Android operating system and the iOS operating system.

Adobe Photoshop is a standardized tool used to develop and edit raster (bitmap) graphics. Using it, the user can draw, paint, process or retouch the photographs, develop designer solutions, create Web graphics, design program interfaces and it also helps in Web page development.

CorelDRAW

CorelDRAW is a vector graphics editor developed and marketed by Corel Corporation of Ottawa, Canada. It is the alternate name of Corel's Graphics Suite, which bundles CorelDraw with a bitmap image editor, Corel PhotoPaint and other graphics related programs. CorelDRAW was originally developed for Microsoft Windows 3 and currently runs on Windows XP, Windows Vista and Windows 7. CorelDRAW is capable of handling multiple pages along with multiple master layers. It is very useful in creating and editing multi-article newsletters, documents, etc. Besides, the other items, such as business cards, invitations, etc., can be designed to their final page size and imposed to the printer's sheet size for cost-effective printing. An additional print and merge feature permits full personalization applications, such as numbered raffle tickets, individual invitations, membership cards, etc.

Corel PhotoPaint

Corel PhotoPaint is a component of the CorelDRAW Graphics Suite and is used to exchange data with other programs in the suite, including Corel CONNECT (Version X5) which enables users to share files between different computer software and the different drives on the user's computer. CorelDRAW and Corel PhotoPaint are copy-paste compatible.

A Corel PhotoPaint is a specific program used for the development and processing of raster (bitmap) graphics. Basically, it comes in a CorelDraw package

NOTES

intended for creation and processing of graphic elements and it can be used as an alternative to Photoshop. The native format of Corel PhotoPaint is .cpt which stores image data as well as information within an image including objects (layers in some raster editors), color profiles, text, transparency and effect filters. The program can open and convert vector formats from CorelDRAW and Adobe Illustrator and can also open other formats including .png, .jpg and .gif files. Corel PhotoPaint is available in English, German, French, Italian, Dutch, Spanish, Brazilian Portuguese, Swedish, Finnish, Polish, Czech, Russian, Hungarian and Turkish.

Macromedia Fireworks

Macromedia Fireworks is a unique program specifically developed for processing and development of raster graphics especially intended for the Internet pages. Generally, it comes in a package with programs, such as Flash and Dreamweaver, and exceptionally harmonizes their functionalities. It is closer to Adobe ImageReady.

Interactive Media: Adobe Flash

Adobe Flash was formerly termed as Macromedia Flash. It is a unique multimedia platform specifically used to add animation, video and interactivity to Web pages. Flash is frequently used for advertisements, games and flash animations for broadcast purposes. Recently, it is considered as a tool for Rich Internet Applications or RIAs. Flash manipulates vector and raster graphics to provide animation of text, drawings and still images. It supports bidirectional streaming of audio and video, and it can also capture user input via mouse, keyboard, microphone and camera. Flash contains an object-oriented language called ActionScript and supports automation through the JavaScript Flash Language (JSFL).

Flash Player is also available to handset manufacturers for smart phones. The Adobe Flash Professional multimedia authoring program is used to create content for the Adobe Engagement Platform, such as Web applications, games and movies, and content for mobile phones and other embedded devices.

1.6 MULTIMEDIA STANDARDS

The term ‘Standards’ refers to the specifications made by the systematic efforts approved by official standardization federations committed to the issue and can sometimes be termed as *official standards*. In some specific cases it is termed as *de facto standards* when it is widely accepted by the industry and/or the public.

The multimedia standards include the following:

- CCITT/ISO (now ITU – T) standards for multimedia include F.700, G.711, G.721, G.722, G.725, H.221, H.242, H.261, H.320, HyTime, IIF, JBIG, JPEG, MHEG, MPEG, ODA, T.80, X.400, G.723, G.726, G.727, G.728, G.764, G.765, H.200, H.241, H.243, T.120.
- Internet standards include IP Multicast, MIME, RTP, ST-2, RFC 741, Xv and mvex.
- W3C standards.
- Proprietary standards are Bento, GIF, QuickTime, RIFF, DVI, MIDI.

There are semi-official standards between official and de facto, such as the Internet and W3C standards.

Image and Video Standards

NOTES

Unprocessed image and video in digital form take up enormous amount of space as compared to text. Usually, a character is one byte and a page can be several hundreds. A color image of the size of a small standard VGA screen occupies 640 ´ 480 ´ 3 bytes, i.e., the number of pixels multiplied by the number of bytes per pixel, one for each of the R, G and B channels. It is extremely important to reduce this size for storage and transmission purposes which can be achieved by coding the image in a compressed way at one end and then decoding it to an uncompressed form at the other end. Thus, the image and video compression standards must be adopted for transmission.

Check Your Progress

14. Fill in the blanks with appropriate words.
    a. To use the multimedia library, a graph is set in case an abstract object encapsulates___________.
    b. A true _____________ platform integrates and combines various multimedia devices and components.
    c. Multimedia ___________ systems streamline data capture by providing interfaces to a range of image and data capture devices.
    d. ____________ is the organisation of information units into connected associations that a user can choose to make.
15. State whether the following statements are true or false.
    a. The most prominent part in a personal computer is the display system that is responsible for graphic display.
    b. Digital representation of sound is governed by bit rate, intensity and flexibility.
    c. The presentation device connects the computer using various wireless technologies, such as Bluetooth connection to produce the presentation online.
    d. The key to successful multimedia production is a seamless integration of multimedia elements for graphic design, content management, production and packaging.

JPEG

JPEG stands for Joint Photographic Experts Group, the original name of the committee that specified the standard. Now the body is officially called ITU – T JTC1/SC2/WG10. JPEG is a *lossy* compression format which means that the decompressed image is usually worse than the original. The compression parameters can be adjusted by the user, trading off file size against output image quality. Recently, a new version called **JPEG2000** has become an official standard. It is based on wavelet transforms and the applications are more ambitious than for the original JPEG.

MPEG

MPEG stands for Moving Pictures Expert Group and is the popular name of the ISO/IEC committee working on digital color video and audio compression, ITU – T JTC1/SC2/WG11, established in 1988. According to MPEG home page, the group is 'in charge of the development of standards for coded representation of digital audio and video'. MPEG has the following versions:

- MPEG-1, a standard for storage and retrieval of moving pictures and audio on storage media. The products, such as Video CD and MP3 are based on it.
- MPEG-2, a standard for digital television. Digital Television set top boxes and DVD are based on it.
- MPEG-4 version 1 and 2, a standard for multimedia applications for the fixed and mobile Web.
- MPEG-7 a content representation standard for multimedia information search, filtering, management and processing.

MHEG

MHEG stands for Multimedia and Hypermedia Information Coding Experts Group, officially called ITU – T JTC1/SC2/WG12. The objective of the standard was to develop a Coded Representation of Multimedia and Hypermedia Information. The standard specifies a coded representation of final form of multimedia/hypermedia information objects to be interchanged as units within or across systems by any means of interchange, from storage devices to telecommunication and broadcast networks.

These objects define the structure of multimedia/hypermedia presentation in a system independent way. These MHEG objects provide functionality for final form representation, support for systems with minimal resources, interactivity and multimedia synchronization, real-time presentation and interchange.

NOTES

1.7 SUMMARY

In this unit, you have learnt that:

- Multimedia refers to a mixture of interactive media or data types, predominantly text, graphics, audio and video that are simultaneously delivered by a computer.
- The multimedia devices and drivers are managed by the [mci] and [drivers] section of the Windows SYSTEM.INI file, that can be added and deleted using the Multimedia Properties control panel.
- The most prominent part in a personal computer is the display system that is responsible for graphic display. The display system may be attached with a PC to display character, picture and video output.
- The aspect ratio of the image is the ratio of the number of X pixels to the number of Y pixels.
- An image in raster scan display is basically composed of a set of dots and lines; lines are displayed by making those dots bright (with desired color) which lie as close as possible to the shortest path between the endpoints of a line.
- The primary component in an electron gun is a cathode (negatively charged) encapsulated by a metal cylinder known as the control grid.
- In raster scan method, the electron beam sweeps the entire screen in the same way you would write a full page text in a notebook, word by word, character by character, from left to right, and from top to bottom.
- In random scan technique, the electron beam is directed straightway to the particular point(s) of the screen where the image is to be produced.
- The various wireless devices and technologies, such as Bluetooth and Wi-Fi connections are required to correlate with presentation device and user interface.
- Bluetooth is used to wireless Personal Area Networks (PANs). It connects and exchanges the information between devices, such as mobile phones, laptops, personal computers, printers, digital cameras and video game consoles via a secure, globally unlicensed short-range radio frequency.
- The Wi-Fi stands for Wireless Fidelity. It is used for wireless devices. Wi-Fi Multimedia (WMM, formerly known as ‘wireless multimedia extensions refer to QoS (Quality of Service) over Wi-Fi.
- When telecommunications access and core networks are concerned, they must be defined clearly in terms of content as well as characteristics. This means that the choices of technology and configurations must be verified in a controlled environment.
- A true multimedia platform integrates and combines various multimedia devices and components.

NOTES

- The key to successful multimedia production is a seamless integration of multimedia elements for graphic design, content management, production and packaging. The whole process of developing a multimedia package is called authoring.
- An authoring system is a collection of software tools that help in various aspects of multimedia production. Multimedia elements are woven together using authoring tools.
- Hypertext is the organisation of information units into connected associations that a user can choose to make. An instance of such an association is called a hyperlink. When a user clicks on such a link, more information on the particular topic is displayed.
- Animation gives visual impact to your multimedia application. In simple terms, it can be defined as an entity moving across the screen. This entity could be a text object or an image. An animation consists of a series of rapidly changing objects, which when blended together gives an illusion of movement.
- Sound is used to set the rhythm or a mood in a package. Speech gives an effect of a language (pronunciation) for instance. Proper usage of sound can make all the difference between an ordinary multimedia presentation and a professional one.
- If pictures can paint a thousand words, then motion pictures can paint a million. Digital video is the most engaging of multimedia venues and is a powerful tool for bringing users close to the real world.
- In computer technology, cross platform or multi-platform refers to the unique characteristic of computer software which enables it in implementing methodologies for inter-operating on several computer platforms.
- In multimedia, the cross platform is the unique capability of an application that helps in accurately performing on a range of computers, operating systems and Web browsers.
- Typically, multimedia on the Web is pictures, sound, music, films, videos and animations. Nowadays, the Web browsers support various multimedia formats because the recent Web pages have embedded multimedia elements.
- Multimedia presentation is of two types, live or recorded. In recorded multimedia presentation, the interaction is possible through a navigation system while in a live multimedia presentation the interaction is done with the help of a presenter or performer. The example of a live multimedia performance is a laser show.
- The term ‘Standards’ refers to the specifications made by the systematic efforts approved by official standardization federations committed to the issue and can sometimes be termed as official standards. In some specific cases it is termed as de facto standards when it is widely accepted by the industry and/or the public.

1.8 ANSWERS TO 'CHECK YOUR PROGRESS'

NOTES

1. Multimedia refers to a mixture of interactive media or data types, predominantly text, graphics, audio and video that are simultaneously delivered by a computer.
2. If the image resolution is more as compared to the inherent resolution of the display device, then the quality of the displayed image gets reduced.
3. Capture devices are used to capture information. Some of the popular capture devices are keyboard, mouse, tactile sensors, etc. For example, a video camera captures visual images still as well as moving in suitable formats. Scanners copy documents or images in image formats, such as jpg, gif, bmp, etc. Video recorder records captured images, still or in motion. Audio microphone captures sound.
4. The PVR or Personal Video Recorder let you record TV programming, and also let you to pause and rewind the live broadcast videos and movies. With this program you get complete PVR functionality to pause the live TV, rewind, fast-forward, and conveniently record your shows for playback on your monitor or TV. It also provides complete media center functionality to access your music, video and photo collections and playback DVDs, etc.
5. Digital representation of sound is governed by three factors. These factors are information in terms of bit rate, complexity and flexibility.
6. The headset is considered as prime presentable devices while using the multimedia files. The leather and metal finish puts engineered longevity and impression to make the desirable sound and view. For example, B&WP5 headphones are virtually leak-proof design that makes it a good choice for entering into the virtual world. Sound and voice, as it often the way, is slightly focussed into the more qualified success.
7. Multimedia authoring tools are software programs that are used for developing a variety of multimedia products.
8. The various types of multimedia software are as follows:
    - Multimedia authoring tools
    - Multimedia tools for the Web
    - Multimedia presentation software
9. Some of the known multimedia platforms are Apple Macintosh and IBM Compatible PC.
10. The term 'authoring,' is used to create content that serve a particular purpose. So, a person who develops a multimedia presentation, authors that multimedia presentation.
11. Hypertext is the organisation of information units into connected associations that a user can choose to make. An instance of such an association is called a hyperlink.
12. Cross platform or multi-platform refers to the unique characteristic of computer software which enables it in implementing methodologies for inter-operating on several computer platforms.

NOTES

13. In recorded multimedia presentation, the interaction is possible through a navigation system while in a live multimedia presentation the interaction is done with the help of a presenter or performer.
14. (a) Multimedia filters, (b) Multimedia, (c) Authoring, (d) Hypertext.
15. (a) True, (b) False, (c) True, (d) True.

1.9 QUESTIONS AND EXERCISES

Short-Answer Questions

1. Write the various hardware components of multimedia system.
2. What are multimedia devices?
3. What do you mean by presentation devices?
4. Write short note on:
   (a) Bluetooth
   (b) Wi-Fi
5. What are authoring tools?
6. What role does cross platform compatibility plays in multimedia?

Long-Answer Questions

1. Explain the need of multimedia.
2. What are different types of display devices? Explain them in detail.
3. Describe various elements of multimedia.
4. Explain all classes of multimedia devices.
5. What are different types of commercial tools?
6. What do you mean by multimedia standards? Give some examples.

UNIT 2 MEDIA TYPES

NOTES

Structure

2.0 INTRODUCTION

In this unit, you will learn about various media types. Text is a fundamental building block in a multimedia system. It is one of the most widely used and flexible means of presenting information and conveying ideas in a multimedia environment. In order to represent text in a digital form, each character of a particular language has to be related to a specific bit pattern. Hypertext is a special type of formatted text. HyperText Markup Language (HTML) is a document-layout and hyperlink-specification language that is used to create hypertext documents and web pages.

You will also learn about compressed data. It occupies less space for storage and also takes less time for communication. The data may be text, image, audio, video or animation objects. There are fundamentally two types of data compression—lossy and lossless. Huffman coding functions by analysing the relative frequency of occurrence of different characters in a text file.

You will also learn about images. An image is the representation of an object or a two- or three-dimensional scene on a planar region (spatial representation). Images can be generated and stored in a personal computer in two typically different ways, such as vector graphics and bitmapped.

Masking occurs when one sound prevents us from hearing a second sound. When you hear a loud sound and a soft sound simultaneously, your ears receive both the sound signals but our brain ignores the soft one and as a result you cannot 'hear' the soft sound. This phenomenon is known as masking. MIDI is a standardized protocol or procedure set. Manufacturers of musical instruments, computers and computer software now routinely adopt MIDI protocol.

Finally, you will learn about extended images and digital ink. Extended images, such as route panoramas, scene tunnels, panoramic views and spherical views are

acquired in an urban area and associated with geospatial locations. To generate a scanning path based on visibility, image properties, and importance of scenes a 3D LIDAR elevation map is used.

2.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Discuss about types of text used in multimedia in different variations
- Identify different font types that facilitates digital text processing
- Describe the various color models and the steps involved in image processing
- Explain the digital image interface standards
- Learn about the basics of digital audio processing
- Explain the significance of and video speech recognition
- Understand the basics of extended images
- Define digital ink

2.2 NON-TEMPORAL MEDIA TYPE

Non-temporal media in also known as a static media. It has the same representation regardless of time. The following sub-section will discuss various non-temporal media types.

2.2.1 Text



Text: A fundamental building block in a multimedia system

Text is a fundamental building block in a multimedia system. It is one of the most widely used and flexible means of presenting information and conveying ideas in a multimedia environment. It is used in a multimedia environment to communicate something in a written language. It can be in English, Hindi, Bengali, Russian, Chinese, etc,.

Now each language has its fundamental units—called *characters*. The characters form a set of alphabet to be used in written communications.

An *alphabet* is a complete standardized set of letters—the basic written symbols to communicate in a particular language. For instance, the English alphabets include **a, b, c, d, …, z**; the punctuation marks (comma, full stop, exclamation, etc.); the digits **0, 1, 2, 3, … 9** and the common symbols (such as + – =); etc. For certain languages, such as Chinese and Japanese, alphabets consist of a set of symbols or characters where each character or symbol represents a whole word or concept. The Japanese Kanji alphabets contain at least 2111 symbols.

Text is used in multimedia in three different variations (i) unformatted or plain text, (ii) formatted text and (iii) hypertext.

Types of Text

An overview of all the three types of text follows:

Plain or Unformatted Text

The plain or unformatted text is the most elementary fixed size character sets. The .txt file created using notepad is an example of plain text. In order to represent text in a digital form, each character of a particular language has to be related to a specific bit pattern. Or in other words, you can store each character of a particular alphabet digitally, if you can map each character with a particular value (to be stored as a Bit pattern).

You can understand that to store the 26 letters of the alphabet for both upper and lower-case letters, along with the punctuation marks and the ten digits (0 to 9) only 26 number of unique *code values* (also called *code points*) is required, whereas for the 2111 characters of the Japanese Kanji alphabet, 2111 number of unique code values is required. Thus, the code points required for representing the English alphabet is significantly less than that required to store the characters in the Kanji alphabet.
It is imperative that in order to effectively share the same textual content between different makes of computers and transfer texts over networks among computers from different manufactures, standardization of the character sets for each alphabet is a must.

Formatted Text

Formatted text: In a formatted text, control characters manage the appearance of the text

In a **formatted text**, control characters manage the appearance of the text. As a result, you can make a string of text appear in any combination of bold, underlined, italic, paragraphed and tabulated style. Such formatting options are available in most text processing software, for example, MS Word, and other publishing software. Using these features a whole document may be formatted in a specific style for the paragraphs, sections and chapters. At the same time, a single character or a word in a document can also be formatted.

The control characters used in the application software may vary. So the appearance of a document created using MS Word may look different in an HTML document and vice versa. However, the control codes are getting more and more standardized so that the text formats, indentations, tabulations, etc., remain the same when you print the document or view them using another word processor.

Remember, in a formatted text (also called rich text) apart from using character strings of different sizes, shapes and styles, you can also use tables, images and graphics at suitable locations.

Hypertext and HTML

Hypertext is a special type of formatted text. In the context of text being used as the fundamental building block of multimedia applications, the powerful processing capabilities of a computer can be applied to make the text more interactive and organize the content in non-sequential way. By positioning the mouse pointer on a portion of a text (a word or even a paragraph on the screen called *anchor*) and then clicking, you may jump to the linked destination and display multimedia information (text, image, video, etc.) in the same screen or on another screen. *Hypertext* is a special text format that is used to link multimedia information in a non-sequential way. In other words, instead of displaying information in a hierarchical way (like a long article written

NOTES

HyperText Markup Language (HTML): A document-layout and hyperlink-specification language that is used to create hypertext documents and Web pages

on a piece of paper), you can organize them in logical blocks and create information path or links to navigate. This is what hypertext is. It allows learners to browse through the text material in a way that suits them; there is no predetermined order as such in which the text is to be read. One cannot do this type of nonlinear and associative navigation in a sequentially organized book.

The meaning of the word ‘hyper’ is something close to ‘extra’ or ‘beyond’. That is, there is something more behind the text you see. It may be connected to another related text material, which again may have links to some new set of words or sections or even whole new documents, and so on. Hypertexts are marked in some visible way (colored or underlined), to differentiate them from the ordinary text. On positioning the cursor over a hypertext, it changes to a pointing finger. If you click the text you can jump directly to additional available information through a predefined cross-reference or link. A particular piece of information in a hypertext system can, thus, be approached from a variety of reference points or nodes. For example, you must have experienced navigating through the help-module of Microsoft Word or Windows or any other good package. All the terminologies in a page that need further explanation or illustrations are hyperlinked to relevant help pages or sections.

HyperText Markup Language **(HTML)** is a document-layout and hyperlink-specification language that is used to create hypertext documents and Web pages. Almost all documents viewed on the World Wide Web (WWW) are HTML documents. It tells how to display the contents of the document including text, images and other supported media.

Basically, HTML files are just plain ASCII text files that can be created in any standard word processors even in Windows Notepad. Such files contain two things—the normal textual content and the markup ‘tags’. These tags are HTML instructions written within ‘<’, ‘>’ symbols specifying the presentation format (such as size, font, color, location, etc.) of the textual content. The markup tags are usually paired with an ending tag starting with a slash (</[tag]>). For example, any text between the tags and will be displayed in bold by the browser. Web browsers, as their name suggests, are used for browsing through web pages on the Internet Websites. The browsers can edit, save, read and display HTML files. Netscape’s Navigator and Microsoft’s Internet Explorer are the two top ranking Web browsers. Both offer a core set of features conforming to HTML so that text, images and links can be handled. Tags can be used to establish hyperlinks to documents, image files, music files, Java applets, etc., from within the document. If the HTML file contains a <a href> tag, the browser knows that what follows describes a hyperlink to another document. Tags are often followed by a list of tag attributes. For example the <img> tag, which embeds an image in the document, can be used as <img src = ‘images/sample_pic.gif’ border = “0” width = “30” height = “27”>. src, border, width, height are all attributes of the <img> tag with attribute values within’. The <HTML> and the </HTML> are the first and last lines of a HTML file. Files without the <HTML> tag can be misinterpreted as text only file, and the markup tags as mere text.

HTML presumes a Document Type Definition (DTD), which specifies valid tag names, attributes and their syntax. However, once you understand their properties and uses, coding or marking up a document is quite simple. HTML translators are built into many words processing software (such as MS Word, WordPerfect, etc.). So you can save a word processed document with its text styles and layout converted to HTML tags for headers, bolding, underlining, indenting, and so on. These works well for simple text documents but the real power of HTML can be exploited only in dedicated WYSIWYG (What You See Is What You Get) HTML editors, such as *Microsoft's FrontPage, Adobe's Pagemaker* or *HoTMetal Pro from Soft Quad*. The *Netscape Communicator* editor offers a point-and-click interface for inserting valid HTML tags; elements; Java applets; JavaScript, and supports in-line plug-ins, such as Acrobat; Shockwave; RealAudio and others, so that multimedia can be incorporated in your HTML document.

NOTES

Font: A font is a collection of characters of a specific style and size of a particular typeface

Font

A `font` is a collection of characters of a specific style and size of a particular typeface. For example, Times New Roman is a typeface and you may choose different fonts (having specific style and size) from within it:

Times New Roman 14 point *Italic*—one font.

Times New Roman 14 point **Bold**—another font.

Times New Roman 12 point ***Bold and Italic***—yet another font.

Similarly, Arial 12 point Normal is a type of font.

And, Arial 12 point Bold is another font

Typefaces are the shapes or graphic representations of the characters, numbers or special characters that are stored internally in the computer as bits. So, do not confuse between font and typeface. Some of the common typefaces are Times, Times New Roman, Arial, Courier, Sans Serif, etc. You may choose different fonts out of these typefaces.

The three basic font styles are—Regular, **Bold** and *Italic*. Other widely available font styles are as follows:

Underline, Super$^{\text{script}}$, Sub$_{\text{script}}$, Shadow, Emboss, Engrave, ~~Strikethrough~~, **Colour**, etc.

Font style is usually measured in *Point*, where 1 Point = 1/72th of an inch (or 0.0138 inch). Thus, a 72 point character will be printed or displayed on the screen (in 1:1 scale) exactly 1 inch high, a 36 point character 0.5 inch high, and so on.

Now let us get acquainted with a few terminologies that are lent from the traditional printing press and are widely used in the digital text processing:

NOTES



Fig. 2.1 *An Example of a Typeface*

Ascender**:** This is the *upstroke* or portion of the character that goes above the mean line or upper level of the normal lower-case character (such as the upper portion of the character 'E' in Figure 2.1).

Descender**:** This is the down stroke or bottom portion of the character below the baseline (as in the letter 'p', given in Figure 2.1).

Cap Height**:** The total height (in *point*) of the capital letters in the font family. (see Figure 2.1).

x-Height**:** Considered as the basis of measurement of the lower-case letters (the letter x is chosen, as it neither has the ascender nor descender).

Leading**:** (pronounced as 'Ledding') used to measure the vertical distance between two lines of text. This has originated from the days of conventional printing press where two lines of texts were separated by thin strips of lead.

Kerning**:** The horizontal spacing between characters. In many word processing software the kerning can be increased or decreased to make the letters in the words more spread out or compact.

S a m p l e T e x t

Font: Arial 48 pt, Bold – with kerning

Sample Text

Font: Arial 48 pt, Bold – without kerning

Example of Kerning

Example of Kerning

There are basically two types of typefaces—and thus two types of fonts—*Serif* and *Sans serif*.

Serif Font**:** There are some typefaces that have decorative features or flags added at the end of the strokes. The fonts using these typefaces are called serif fonts. Examples

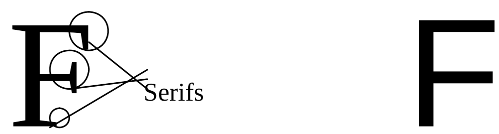
are Times New Roman, Courier New, Monotype Corsiva, etc. Serif fonts are generally used for body text as the serifs make reading easier.

***Sans-Serif Font*:** These fonts use typefaces that are without (or sans) any decorative features or flags at the end of the strokes. Examples are Arial, Verdana, Impact, Tahoma, etc. Sans-serif fonts are generally used for headlines or bold statements.

NOTES

Some Sans-Serif Fonts
Arial
Verdana
Impact
Tahoma

Some Serif Fonts
Times New Roman
Courier New
Monotype Corsiva
Bookman Old Style



Serifs are decorative flags   The same letter 'F' in Arial

Font Types: PostScript, TrueType and Bitmap Fonts

***PostScript Font*:** In the early days of text processing and DTP, Adobe introduced a method of printing and displaying text using special software called Adobe PostScript. It was a licensed software and was available under license in proprietary printers and operating systems. Later, Adobe introduced Adobe Type Manager that could display postscript fonts on both Macintosh and Windows monitors. Multimedia developers now rarely use it.

***TrueType Font*:** Apple and Microsoft jointly developed a font technology called TrueType font. It could print smoothly on paper and display clearly on even low-resolution monitors. This effectively freed both Apple and Microsoft from paying the license fee to Adobe for using the PostScript Font in their operating systems. The TrueType font, unlike the PostScript font, does not need any special software to display.

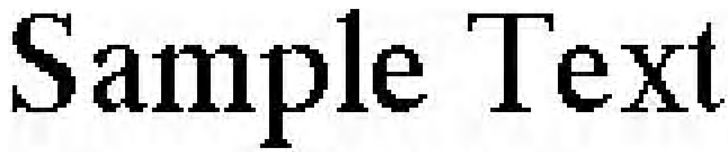
***Bitmap Font*:** Unlike the PostScript and TrueType fonts, the bitmap font actually uses the images of each character. For each letter typed, a bitmap image representation of the letter is inserted. Thus, it requires a lot of memory. However, the quality of the output is constant for a particular font style and size.

***Bitmapped and Vector Fonts*:** Fonts can either be stored as bitmapped or vector information. *Bitmap fonts* use one bitmap for each size of a particular typeface. As the number of fonts (i.e., different typefaces and their sizes) increase the bitmap information becomes huge thus increasing the file size and also requiring a lot of memory.

The *vector fonts* draw the characters by using vector drawing primitives using mathematical functions, thus requiring considerable smaller size than the bitmap fonts. Also the fonts can be drawn in any size without restrictions as generalized vector drawing primitives are used. The PostScript and TrueType fonts are examples of vector fonts.

NOTES

For example: Type some text in a Notepad and change the text size to 48 points. First use MS-Sarif Font, which is a bitmap font, and see how the letters are pixelated. Next, change the typeface to Times New Roman—see how the letters look smooth now.



The text typed in MS-Serif Font (bitmap font) in a Notepad
See how the letters look jagged.



The font is changed to Times New Roman (TrueType font).
Observe how the letters now look smooth.

Font Mapping**:** Some fonts available in the multimedia developer's machine may not be available in the user's machine. In such a case, a default font will be automatically used, specifying which font will be substituted is known as *font mapping*.

Text Compression

You must be aware that while processing digital multimedia data, smaller files are desirable for faster data communication as well as economical storage of data. Data compression brings down the size of a digital data file so that the preceding objectives are achieved. Compressed data occupies less space for storage and also takes less time for communication. The data may be text, image, audio, video or animation objects. There are fundamentally two types of data compression—lossy and lossless. As the name suggests, lossy data compression results in a compressed data file, which when decompressed may not be recovered exactly as it was before it was compressed. Lossy compression techniques are widely applied for compression of image, audio and video data as the lossy data compression algorithms utilize the limitations of our eyes and ears and discard portions of the image or audio signals to achieve compression and bring down the file size. For example, lossy compression is applied in JPEG image format and MP3 files. Moreover, the image, audio or video data are inherently analog in nature and when digitized some data loss invariably creeps in during sampling and quantization. However, in the context of textual data, lossy compression techniques should not be used, as one cannot afford to lose a few characters in a text file, which may make the text meaningless. Imagine applying lossy compression to a computer program source code (which is a text file), and the source code is changed after decompression due to data loss. Lossless data compression usually works by identifying repeated patterns in a data and encoding those patterns efficiently. In other words, lossless data compression reduces the redundancy of the patterns in a message. Lossless data compression is ideal for text. However, it is also used for other media, such as

image, for example, the GIF and PNG image file formats use lossless compression algorithms only. You will learn the different lossless compression algorithms that can be applied to text compression.

Huffman Coding

Huffman coding functions by analysing the relative frequency of occurrence of different characters in a text file. The characters in the text file that have the highest frequency of occurrence are assigned the shortest encoding with the fewest bits. Characters with lower frequencies get assigned longer encoding with more bits. Thus, compression is achieved by overall saving in the total number of bits.

Lempel-Ziv (LZ) Coding

In Huffman coding the frequency of each character in the text file is analysed for the encoding operation. In the *Lempel-Ziv* method, instead of using a single character, a repetitive string of characters is encoded and a table is maintained by both the encoder and the decoder when the character string is encountered repetitively, instead of using the ASCII characters, the encoder points to the index number against which the string is stored in the table. During decompression the decoder converts the index number to the original string as per the table. For example, consider the text, '*she told me you told her I told you not to tell her*'. Here, the strings 'told', 'you' and 'her' can be represented only once and subsequently pointed to by all later calls to those strings.

File Formats

As you have already seen, text can be used in a wide variety of applications, such as computer program source code, log file, mail message, formatted or unformatted text to name a few. In the context of text as a multimedia object, you will note down a few text file formats here.

Unformatted Text (TXT)

The **unformatted text** documents can be created by Windows Notepad or any standard program editor. The data can be encoded in ASCII or Unicode (UTF-8 or UTF-16).

Formatted Text (DOC)

The formatted text document created using Microsoft Word or WordPad packages have by default the .doc extension. This format is very common and has a rich set of formatting features. It also supports images and graphics. Most open source word processing software these days supports this format.

Portable Document Format (PDF)

This format was developed by Adobe Systems for cross platform file exchange. The format supports image and graphics. It requires proprietary software to create (Adobe Acrobat Writer). However, the Adobe Acrobat Reader (PDF) is available free of cost.

Portable document format: This format was developed by Adobe Systems for cross platform file exchange. The format supports image and graphics

Rich Text Format (RTF)

This format was developed in 1987 by Microsoft for exchange of formatted text across different platforms. It can be edited by MS Word or WordPad. The control characters are editable.

NOTES

Image: It is the representation of an object or a two- or three-dimensional scene on a planar region (spatial representation)

2.2.2 Image

An **image** is a very important component of digital multimedia. It is the representation of an object or a two- or three-dimensional scene on a planar region (spatial representation). It can be a photograph, a map or an analog video signal (a video footage is nothing but moving images and audio). In the context of computer graphics, the image is always in digital format. It is, therefore, very important to understand how digital image is created, stored, edited and transmitted, as it is the fundamental building block for both graphics and video.

A digital image can be considered as a set of picture elements (pixels). These pixels are like the tiny dots or pigments on a photograph printout arranged in rows and columns that makes up the image. Each pixel corresponds to a color value at a particular portion of the image.

Image Types

Images can be generated and stored in a personal computer in two typically different ways. One is called vector graphics and the other is referred to as bitmapped.

A piece of *vector art* is a file that contains descriptions of how to generate the image but not the actual image itself. A vector graphics program (generically called drawing program) creates a sequential list of graphic commands to draw lines, curves, text, etc., with associated parameters, such as screen location, size, color, rotation angle, width, style, etc. This type of list-file is often referred to as a *display list/file*. Such a file must be rasterized before it can be presented as an actual image on screen.

While a vector graphic is edited, the properties of the lines and curves, which explain its shape, are changed. Without altering the quality of its form, one can reshape, resize, move and modify the color of a vector graphic. Being resolution-independent, vector graphics may be displayed on output devices of varying resolutions without losing any quality.

Mostly, the CAD packages, business packages for drawing charts and graphs and some DTP package, such as Corel Draw, etc., use vector files of specific formats. Some of the vector file formats commonly used in IT industries are:

- Postscript file.
- Computer Graphics Metafile (*.CGM).
- Windows Metafile (*.WMF).
- Hewlett Packard Graphics Language or HPGL (*.PLO).
- Data Exchange Format (*.DXF).

PostScript files, developed by Adobe, are generated by DTP packages and authoring systems while WMF was developed by Microsoft, and it is an excellent format for image interchange between Windows applications. HPGL is an interpreted vector description language meant for plotters, and DXF is the most widely accepted format for interchange of engineering graphics data between different CAD packages, such as AutoCAD, etc.

Another type of vector image specifies the content in full three-dimensional form. Here the objects are not what get drawn in the image. Instead a view of those

objects is drawn. This substantially more complex process is called three-dimensional rendering.

A *bitmapped image*, in contrast, has in the file the actual pixel image data. That is, it simply holds the color number for each dot or pixel in an image. The size of such files depends on the image size and how many colors are to be used per pixel, i.e., the color depth. For a 256-color range and standard VGA (640 × 480 pixels) full screen display the size of a bitmap file is 640 × 480 × 8 bits, i.e., 307200 × 8 bits or 307200 bytes or 300 KB. This is because for each pixel 8 bits are required for storing color value anything up to 256. However true-color images (24 bit, more than 16 million color) provide the highest quality, and they are the best way to represent photographs on computer screen. Out of the 24 bits per pixel in high quality true-color images, 8 bits are used to describe intensities of each of the three basic color signals—red, green and blue of *RGB color model*. White color is displayed when all RGB signals are at full intensity and black occurs when there is no signal.

NOTES

As soon as a bitmap graphic is edited, the pixels are altered but not the lines and curves. The former are resolution-dependent as the data unfolding the image is set to a grid of a particular size. However, editing a bitmap graphic alters the quality of its appearance. For instance, resizing a bitmap graphic make the edges of the image ragged because pixels are redistributed within the grid. Showing a bitmap graphic on an output device that has a lower-resolution than the image also degrades the quality of its appearance.



***Left*:** On the left, the vector image of a leaf is shown by points through which lines and curves pass making the shape of the outline of the leaf. The color of the leaf is determined by the color of the outline and the area enclosed by the outline.

***Right*:** The bitmapped-image of a leaf is described by the specific location and color value of each pixel, creating a more realistic image.

Color Models

The various color models are discussed in the following sections.

Light is an electromagnetic wave. The human eye is able to 'see' only a very small part of the total range of electromagnetic radiation called the *visible light*. The wavelengths light waves visible to a normal human eye is roughly 400 to 700 nanometres (1 nm = 10^{-9} metres). You cannot see light waves with wavelength above 700 nm (infrared light) or below 400 nm (ultraviolet light). The visible light range in the perspective of the total range of electromagnetic radiation is shown in Figure 2.2.

NOTES

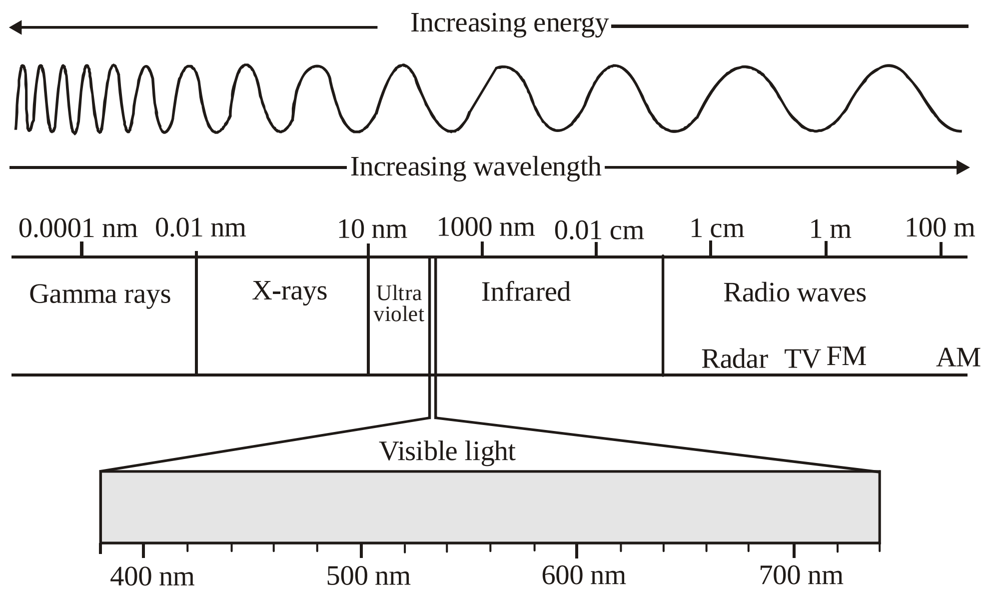


Fig. 2.2 *Visible Light Range*

Within this narrow frequency band of visible light, you can perceive different colors for different wavelengths of the light wave. When the light waves fall upon the color receptors of our eyes, our brain translates the contact between the waves and the eyes as color perception. So the perception of color is a complex physical and psychological phenomenon.

The nature around us is colorful. Yet, you can hardly ever see around a pure color of a single wavelength. A mixture of wavelengths generally creates the colors you see around us. The red color of a cricket ball may look like pure red to us, but a spectrometer analysis will reveal that it is actually a combination of wavelengths (having wavelengths near the 700 nm range). A spectrometer is used to break up a color into its component wavelengths.

Light waves having a single wavelength are called monochromatic light, and the colors produced by a visible light of a single wavelength are pure spectral colors. The light produced by a laser torch is monochromatic, having a single wavelength.

Two well-known color models, namely the HSV and HLS color models, adopt this approach. Another way to look at color is as a combination of three primary colors. This color model is also called device color model because output devices, such as the Cathode Ray Tube (CRT) or the Liquid Crystal Display (LCD) monitors utilize the Red Green Blue (RGB) color model and color printers utilize the Cyan Magenta Yellow (CMY) color model. Different color models have advantages in different situations, and one cannot identify a single color model as the best. The different color models are briefly discussed as follows:

RGB Color Model

One approach to create a broad range of colors is by suitable combinations of three primary colors. Three colors are primary with respect to one another if none of them can be created by any combination of the remaining two. For instance red, green and blue are the three primary colors, as you cannot create red by the combination of green and blue, green by combination of blue and red, and so on.

The red, green and blue (RGB color model) is chosen because the cone cells in the human eye that are responsible for sensing color of a light, or the colors receptors, are particularly sensitive to these three hues.

Any color can be defined within the color gamut of the RGB color model by combining suitable amounts of red green and blue light energy respectively. In other words:

$$C = rR + gG + bB$$

where *r*, *g*, and *b* are the relative amounts of red, green and blue color and *C* is the resultant color.

CMY Color Model

The **CMY color model** is another model where the primary colors chosen are Cyan (***C***), Magenta (***M***) and Yellow (***Y***). In the *CMY color model* a color is divided into three primaries—C, M and Y, using a subtractive rather than an additive color creation process.

The CMY model is widely used in professional four color printing processes. The color printer is programmed to combine different amounts of cyan, magenta and yellow inks to create a color. Since in the subtractive model the addition of C, M and Y make black color, ideally you should have used maximum amount of cyan, magenta, and yellow ink to produce black. However, in reality it produces not black but a muddy brown color. That is why, in actual practice, a fourth component—pure black ink is used in the professional four-color printing press and in all the standard color laserjet and inkjet printers.

HSV and HLS Color Models

You have seen how color is represented by three primary color components. The main disadvantage with primary color combination (RGB or CMY) is that it not intuitive. That is, you cannot intuitively guess what amount of r, g and b or c, m and y would combine to produce a color, such as brown or orange (say). The Hue, Saturation and Value (HSV) and Hue, Lightness and Saturation (HLS) models have been developed on a more intuitive approach. In this method, it is possible to describe a color in terms of its *hue* (i.e., the perceptual similarity with essential color, such as red, green, blue and yellow), its *lightness* (*luminance* or *value* or *brightness*), and its *saturation* (i.e., the purity of the color). Both the HSV (or HSB) color model as well as the HLS model represent color in this way.

The HLS color model is essentially the same. You can create the HLS color space from the HSV color space and take a mirror image of the hexacone to get a double cone. The hue and saturation are represented as in HSV model, but the lightness parameter varies from 0 at the black point to 1 at the white point at the opposite end of the double cone.

Steps in Image Processing

Digital image processing refers to processing of digital images by means of a digital computer where the input is an image and the output is another image or a set of

NOTES

CMY color model: Another model where the primary colors chosen are Cyan (C), magenta (M) and Yellow (Y)

NOTES

characteristics or parameters acquired from the image. The basic steps in image processing as applicable to digital multimedia are:

- **Image Acquisition:** Image may be acquired from analog source, such as a conventional photograph or a frame of a video clipping, or it may be available in digital form like a photograph taken with a digital camera. Generally, digital camera and scanners are the input devices for digital input.
- **Image Editing:** As the name suggests, image editing in the context of digital multimedia involves the following processes all or some of which may be applied on the image that has been acquired in digital form:
  - o ***Enhancement*** to bring out the obscure part of an image or highlight certain portion. Image enhancement is a very subjective topic and demands technical as well as artistic acumen.
  - o ***Image Restoration*** is an area where unlike the image enhancement, which is subjective, the quality of the image is improved based on mathematical parameters of the image data and rectify any degradation already existing in the image.
  - o ***Compression*** as the name suggests, compression deals with techniques for reducing the storage requirement or the bandwidth required to transmit it.
- **Image Output:** Once the image has been acquired in digital form and edited or processed, the edited image is generally out put using a standard output device, such as a computer monitor or a printer. Or else, the image may be utilized in another multimedia object like a video or an animation.

Interface Standards

With the popularity of digital cameras/webcam and scanners increasing and their cost coming down, huge number of makes and models of these data acquisition devices have flooded the market. This has led to the necessity for standardizing the interface between the data acquisition devices and the computer in an easy plug-and-play way. You will learn here two main digital image interface standards—TWAIN and ISIS standards.

TWAIN

TWAIN is an image capture API, developed by consortium of Hewlett-Packard, Kodak, Aldus, Logitech and Caere for Microsoft Windows and Apple Macintosh operating systems. The API uses a four-layer protocol (device layer, acquisition layer, protocol layer and application layer) for connecting TWAIN compliant devices with TWAIN compliant applications, mostly through USB interface. TWAIN permits software applications to work with image acquisition machines without knowing anything about the machine itself. If a machine is TWAIN compliant and a software application is TWAIN compliant, both should work together regardless of whether the software was put together with the device driver of the image acquisition device when it was purchased. Also, it is possible to connect multiple TWAIN compliant image acquisition devices to a PC simultaneously.

Image and Scanner Interface Specification

The Image and Scanner Interface Specification (ISIS) has more functions than TWAIN and is mainly used with the SCSI-2 interface. While TWAIN is developed and

maintained by TWAIN Working Group – a non-profit organization, the ISIS is developed and maintained by a company—M/s Pixel Translations.

Specifications of Digital Images

You have already learned that digital images can be created in three basic ways: raster graphics or *bitmapping*, *vector graphics* and *procedural modelling*. *Bitmap images* are created as a two-dimensional spatial image by storing pixel information along rows and columns of a pixel grid. *Bitmaps* are generally created by scanners, digital cameras and paint programs, such as Corel Paint Shop Pro®, (image processing programs) Adobe Photoshop®, etc. Vector graphic images, on the other hand, are created using graphics primitives, such as lines, arcs, etc., governed by mathematical equations describing the shapes and colors are applied to those primitive shapes. Examples of vector graphics programs are Adobe Illustrator® and CorelDraw®, Autodesk AutoCAD®, etc. You can also draw digital images by means of a computer program using mathematical functions, control logic and often some recursive procedure, which is called procedural modelling. It is also known as *algorithmic art*. Fractals are examples of procedural *art*. Here you will learn about the characteristics or the specification that affect the quality of digital image the resolution, color depth, color palette, etc.

Resolution

In graphics, each pixel represents one sample of a portion of the image area. Finally, the image is divided into pixel information representing a uniform two-dimensional grid. The number of pixels across a row or down a column corresponds to the number of samples taken to represent the spatial resolution to which the picture has been sampled. The more are the rows or columns, the more are number of pixels and the finer is the spatial resolution at the cost of increasing file size. Resolution gives an idea of the clearness or detail, and can refer either to an image file or the device, such as a monitor, etc., used to display it. Image file resolution is expressed as a ratio, such as 800 × 600; a similar matrix, 1024 × 768, for example, is used to characterize monitor displays. Print resolution is generally expressed in terms of dots per inch (dpi).

Image Size

Image size is the physical dimensions of an image when it is printed out or displayed on a computer screen. It is normally expressed in inches or centimetres. In some image editing software (for example, Adobe Photoshop®) image size can also be expressed in pixels. In that case the pixel resolution (in ppi) is important. Anyway, image size depends on both the pixel dimensions and resolution:

Image size: The physical dimensions of an image when it is printed out or displayed on a computer screen

If an image (of pixel dimension $w \times h$) is printed using a printer set at resolution r dpi, the image size of the printout ($a \times b$ say) will be:

$$a = w/r$$

$$b = h/r$$

For example, an 800 × 600 pixel image, if printed at 200 ppi, will give a 43 × 33 image.

NOTES

Color Depth

In digital system, color depth or pixel depth refers to the number of bits associated with each pixel in a bitmap. Using one color bit per pixel only monochrome image would be produced. An image whose color data is represented by only one pixel has a color depth of one pixel as in a system with a monochrome monitor. To compensate for the limitations the developers can use a process called dithering and can create additional shades from an existing palette by varying the density and patterns of the dots. In color displays, it creates colors and patterns by mixing and varying the colors. Dithering can create a wide variety of patterns for use in backgrounds, fields and shading as well as for creating half tones. When you represent the pixel depth in 16 bits you have 32,000 colors to choose from. 5 bits of data characterize each red, Green and Blue (RGB) signal with one bit of color used for the overlay of the text or other graphics on an image.

Examples of color depth are shown in Table 2.1.

Table 2.1 *Color Depth*

Colour Depth	No. of Colours	Colour Mode
1 bit colour	2	Indexed Colour
4 bit colour	16	Indexed Colour
8 bit colour	256	Indexed Colour
24 bit colour	16,777,216	True Colour

Color Management Systems (CMS)

Different graphics application programs and hardware devices, such as Adobe Photoshop or ImageReady, as well as the scanners, monitors, printers, etc., each have their own color spaces and basic color settings. A Color Management System (CMS) is a collection of software tools designed to look after the different color capabilities of scanners, monitors, printers, image-setters and printing devices to ensure consistent color throughout the process of print production. In other words, the colors displayed on the computer screen should be represented as accurately as possible in the final output. Also, colors should be displayed consistently across different applications, monitors and operating systems.

A color management system serves as a translator and communicates the color settings from one device or software program to another. It also communicates the assumptions about the primary colors used, the color spaces, as well as the mapping from color values to physical representations in pixels or ink from one device to another.

The process of color management involves five steps:

- Calibrating the computer monitor.
- Describing or characterizing the monitor's color profile.
- Creating the color profile of a particular image including the choice of the color model.
- Saving the color profile information with the image.

- Reproducing the image's color on another device or application program based on the source and destination profiles.

File Formats

NOTES

Utilities, such as Windows Paintbrush or Paint Shop Pro, etc., generate *bitmapped* image files in BMP format or in the more efficient PCX format.

Static bitmapped images are often compressed to reduce the file sizes and thus, to save some disk space and shorten the time it takes to transfer those files over a communication link. The most common compressed file formats are:

- Graphics Image Format (*.GIF).
- Tagged Information File Format TIFF (*.TIF).
- Joint Photographic Experts Group JPEG (*.JPG).
- Windows Bitmap (*.BMP) and Windows Device Independent Bitmap (*.DIB)

GIF, TIF, DIB and PCX files are compressed in *lossless* fashion using either *RLE*[*] or *LZW*[†] compression algorithm. That is, only truly redundant bits are squeezed out, and they all can be returned exactly as they were when the file is decompressed. None of the original images data is deleted in the compression process.

JPEG, on the other hand, is an example of a lossy compression. Data from the original image which is deemed to be redundant is thrown away in the compression process. (And as you would expect lossy compression yields smaller compressed files than does lossless compression.) This means that the image resulting from the decompressed files will differ from the originals to some degree. The tricky part of these algorithms is their attempt to lose only 'unimportant' features of the images and people are least likely to notice those absent features while viewing the reconstructed image.

Researches are on for evolving more effective compression techniques. The objective is to reduce the compression as well as decompression cost and time without any significant degradation in image quality but at the same time enhancing the storage savings.

2.2.3 Graphics

The term 'computer graphics' was first used by William Fetter in 1960. Fetter was a graphic designer for Boeing Aircraft Co. The term was actually given to him by Verne Hudson. The demonstration of computer graphics technology led to the development of computer graphics. The projects in computer graphics (like the SAGE and Whirlwind projects) gave an impetus to computer graphics as a discipline by introducing the CRT (cathode ray tube) as a viable display and interaction interface, as also by introducing the light pen as an important graphics input device. The TX-2 computer developed in 1959 by MIT's Lincoln Laboratory further continued the development of digital computers and interactive computer graphics.

A light pen, a display unit, and a bank of switches were the main components connected with the interface on which the first interactive computer graphics system

[*] Run Length Encoding or RLE, in which recurring pixels of same value are stored as a single pixel along with the count of number of times the value is to be repeated.

[†] Lempel-Ziv-Welch or LZW, a proprietary lossless data compression algorithm.

NOTES

was based. The TX-2 architecture was integrated with a number of man-machine interfaces. All that these interfaces needed was a user that would use them to make an on-line computer. A user was able to draw on the computer with a simple cathode ray tube and light pen on the TX-2's console and, the Sketchpad, and with it, the interactive computer graphics was born. On the TX-2 computer in the Lincoln Labs, the scientists' research work made them the 'grandfather' of Graphical User Interfaces (GUIs) and interactive computer graphics. The study and experimentations at the Massachusetts Institute of Technology (MIT) shaped the early computer and computer graphics industries.

Personal Computers (PCs) became more powerful during the late 1970s. These PCs could draw both complex and basic designed and shapes. During the 1980s graphic designers and users realized the significance of the PC, particularly the Macintosh and Commodore Amiga as a fundamental design tool; one that could not only draw more accurately, but also save time compared to other methods. SGI computers, powerful as they were, made three dimensional (3D) computer graphics effective in the late 1980s. Later these were used to create the first fully computer-generated short animations. Macintosh has been one of the most accepted tools for computer graphics in businesses and graphic design studios.

From the 1980s onwards, modern computer systems have frequently used a Graphical User Interface (GUI) to represent data and information with the help of symbols, icons and shortcuts, rather than the text user interface. Graphics are one of the five major key elements in the design of multimedia applications.

Three dimensional graphics became more popular during the 1990s in game designing, multimedia and animations. 'The Quake,' one of the first fully 3D games, was released in 1996. *Toy Story*, the first full-length computer-generated animation film, was commercially released in cinemas worldwide in 1995. Since then, computer graphics have become more truthful and comprehensive, due to more advanced computers and better 3D modelling software applications.

Computer Graphics

Computer graphics: The discipline of producing pictures or images using a computer

Computer graphics is the discipline of producing pictures or images using a computer. It includes creation of model, manipulation and storage of geometric objects, reproduction (an image converted from a scene), transformation (primitive graphics operations), illumination, rasterization, animation, and shading of the image.

Computer graphics are broadly used in such activities as graphics based presentations, paint systems, image processing, simulation, virtual reality and Computer-Aided Design (CAD) and entertainment. From the earliest text character images of a non-graphic mainframe computers to the latest photographic quality images of a high resolution personal computers, from vector displays to raster displays, from two dimensional (2-D) input, to three dimensional (3-D) input, computer graphics has gone through its short, rapid changing history.

Types of Computer Graphics

In general, there are two types of computer graphics: vector (which is composed of paths) and raster (which is composed of pixels). Vector graphics use mathematical relationships between pixels and the paths connecting them to represent an image.

NOTES

Vector graphics are composed of paths. The image in Figure 2.3 (a) represents a bitmap and the image in Figure 2.3 (b) represents a vector graphic. Classically, raster images are more commonly called bitmap images. A bitmap image uses a matrix of individual pixels where each pixel in the image may contain different colors or brightness. Bitmaps consist of pixels. They may be shown at four times of the actual size to overstress the fact that the edges of a bitmap become uneven as it is scaled up.

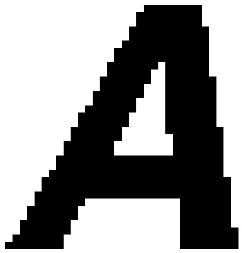


Fig. 2.3 (a) *Bitmap Image*

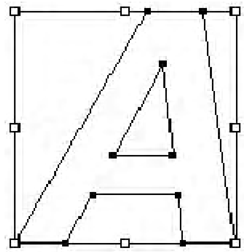


Fig. 2.3 (b) *Vector Graphic*

With the help of various image handling tools and by using point-to-point method rather than by pixels alone, computers can display various fonts and images. When a user scales an image up, the advantage of using a page-description language such as PostScript becomes clear. The more jagged it appears, the larger the user displays a bitmap. On the other hand, a vector image remains even at any size. That is because PostScript and TrueType fonts always appear smooth and they are based on vectors.

One can partially overcome the uneven appearance of bitmap images with the help of 'anti-aliasing.' Anti-aliasing can be defined as the application of frail transitions in the pixels along the edges of images to minimize the uneven effect as shown in Figure 2.4 (a) A scalable vector image always appears smooth as shown in Figure 2.4 (b).



Fig. 2.4 (a) *Anti-Aliased Bitmap Image*



Fig. 2.4 (b) *Smooth Vector Image*

Bitmap images require higher resolutions (large number of rows and columns in the screen) and anti-aliasing for a smooth (even surface) appearance. On the other hand, vector-based graphics are mathematically much more useful and appear smooth at any resolution.

Graphics Primitives

A pixel (also called picture element) can be defined as the smallest piece of information in an image. Pixels are normally arranged in a regular two-dimensional grid (matrix), and are frequently represented using squares, rectangles or dots. Each pixel represents a sample of an original image, where more samples characteristically provide a more precise representation of the original image. The brightness of each pixel is incompatible. In typical color systems (like RGB or CMYK), each pixel has three or four basic color components, such as Red, Green and Blue (RGB system), or Cyan, Magenta, Yellow and Black (CMYK).

The display resolution of a digital television or computer monitor typically refers to the number of distinct pixels in each dimension that can be displayed.

A palette is a given finite set of colors for the management of digital images or a small on-screen graphical element for choosing from a limited set of choices which are not necessarily colors.

Graphics Image File Format

There are several graphic image file formats that are used by most of the graphics systems. The following are the most regularly used formats:

Web Document Images

Web document images can be of two types as follows:

(a) ***Graphics Interchange Format (GIF):*** Images made using this format use a fixed color palette which is limited to only 256 colors. This format downloads small, compressed files quickly from the Web. This format is most suitable for images with solid colors or uniform colour areas, such as illustrations and logos.

(b) ***Joint Photographic Experts Group (JPEG):*** These files are used for photographic (continuous color tone) images, i.e., those images that have a continuous color tone. Unlike the Graphics Interchange Format files, the Joint Photographic Experts Group format takes advantage of the full spectrum of colors available to the display unit. The JPEG format also uses compression for making smaller files and for obtaining faster downloads over the World Wide Web. However, unlike the compression method used in GIF files, the JPEG compression is also 'lossy compression' which means it discards some data in the decompression process. Once a file is saved in the JPEG format some data is lost permanently. But this does not affect the image.

Printed Documents

The following are the two types of printed documents.

(a) ***Encapsulated PostScript (EPS):*** It is an image file format used for both vector graphics and bitmaps. EPS files have a PostScript description of the

Encapsulated PostScript (EPS): It is an image file format used for both vector graphics and bitmaps

graphic data within them. The EPS files are exclusive in that the graphics users use them for bitmap images, vector graphics, type or even entire pages.

(b) ***Tagged-Image File Format (TIFF):*** Such files are used for bitmap format only. The **TIFF** formats are the files that are supported by virtually all graphics applications.

NOTES

TIFF: The TIFF formats are the files that are supported by virtually all graphics applications

2.3 TEMPORAL

The media that has an associated time aspect is called temporal media. Its views changes with respct to time. The temporal of a multimedia system includes the following for its functionality:

2.3.1 Audio

Digital audio applications involves recording sound (with the appropriate sampling rate and sample size), selecting the right type of microphones to suit the specific purpose, using the right type of sound card, RAM, hard disk speed and processor and editing software for recording and editing. Also knowledge and experience are required for compression of audio files, selecting the right file type for storage, applying special effects, and identifying and rectifying the imperfections in recorded audio. In this treatment, you will deal with the concepts underlying digital audio representation and how to apply these concepts in digital audio processing.

Sound or audio is one of the most significant components of digital multimedia. You may have noticed that while all the other multimedia components, such as text, images, graphics and video are sensed through our eyes, sound is essentially perceived through the hearing organ—the ears. However, just like our eyes, our ears, on hearing a sound, send nerve impulses to the brain to stimulate a complex series of psychological and physical response. A haunting melody of the 1960's may set a blissful mood for one person, while it may bring some sad recollections to another and fill the heart with sorrow.

In any quality multimedia production, two things are to be considered while using sound —you have to keep provision for interactivity to control the sound and provision should be there to effectively convey the meaning of the multimedia presentation to the hearing impaired persons or where computers are not available with sound card. You can provide transcriptions and captions to address the second requirement.

Acoustics

Before you delve further into the details of application of audio in multimedia production, let us brush up our understanding about sound wave and its different characteristics. Sound is generated by vibration of matter. The 'matter' or the medium may be solid, liquid or gaseous. It may be a guitar string or the stretched skin of a drum or the air column in a flute or our voice box. As the matter vibrates, pressure variations develop in the medium surrounding it. This alternating high and low pressure travels through the medium (air, water or any solid material) in a wave like motion. When the wave reaches your ear, you hear a sound. Everyday, you hear thousands of sound around you. Yet

they are all different from each other. What make two sounds different? To know this *acoustics* or the branch of physics that studies sound will be briefly discussed.

Nature of Sound Waves

NOTES

You will learn some key terms of acoustics that are often quoted in multimedia world and will be often used in this and in subsequent sections. In this process, you will examine the nature of sound waves and how acoustics is applied in digital multimedia applications.

Frequency

When sound travels through air, alternating high and low pressure are created along the wave path. You can put your hand in front of a loudspeaker in full blast and feel it. The ***frequency*** is the number of alternating high and low pressure oscillations per second at a fixed point occupied by a sound wave. One single oscillatory cycle per second corresponds to 1 Hertz (or Hz in short).

Frequency: The frequency is the number of alternating high and low pressure oscillations per second at a fixed point occupied by a sound wave

The commonly followed abbreviations are:

Hertz → Hz *Kilohertz* → *kHz* *megahertz* → MHz *second* → s	***Key equations*:** 1 Hz = 1 Cycle / s 1 KHz = 1000 Hz 1 MHz = 1,000,000 Hz

Period

The amount of time taken by a wave to complete one cycle is the *period* of the wave. Period and frequency are reciprocals of each other.

If T be the period and f the frequency of a sine wave, then
$T = 1/f$ and $f = 1/T$

Amplitude

When sound travels through a medium the particles are subjected to alternating high and low pressure. The *pressure amplitude* of a sound wave (see Figure 2.5) measures the change in sound pressure inside the wave. In other words, it is the maximum pressure at any point in the sound wave. The pressure amplitude is frequently referred to as sound pressure level and measured in decibels or dBSPL, dBspl or dB (SPL).

Also, sound wave displaces the particles of the vibrating medium and *displacement amplitude* of a sound wave is the maximum displacement of a point on the path of the wave and is measured in units of distance.

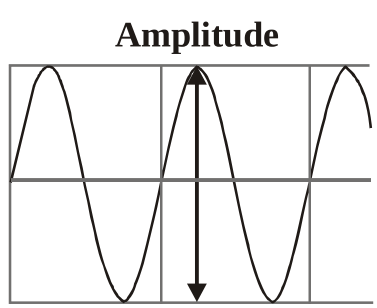


Fig. 2.5 *Amplitude*

Wavelength

The distance between two successive crests is known as the **wavelength**. It is the distance that a wave travels in the time of one oscillatory cycle (see Figure 2.6).

The wavelength of a sound wave of frequency *f* and travelling at speed *c* is indicated by *c/f*. Using this relationship, and the knowledge that the *velocity* of sound in air is about 343 m/sec. we may derive that a 20 kHz (or 20,000 Hertz) sound wave has a wavelength of about 17.15 mm (i.e., 343000/20000). On the contrary, a 20 Hz sound wave has a wavelength of 17150 mm or about 17 m (i.e., 343000/20) and a tone of 343 Hz travelling in air has a wavelength of 1 metre (see Figure 2.6)

NOTES

Wavelength: The distance between two successive crests is known as the wavelength

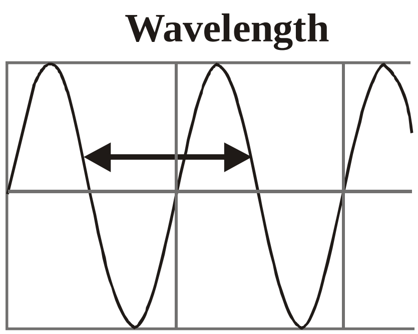


Fig. 2.6 *A Wavelength*

Sound Velocity

The *velocity* of sound in air is about 343 m/sec. The velocity of propagation of sound is dependent on the temperature, type and pressure of the medium through which it propagates. In dry air at 20 °C (68 °F) the speed of sound is about 343 m/s.

Waveform

As the name suggests, waveform is the form or shape of a wave. In acoustics, *waveform* is the shape of a sound wave travelling through a medium. The medium may be gaseous, liquid or solid. In order to study a waveform, you plot the amplitude (pressure or displacement) along the vertical axis and time (or distance) along the horizontal axis. Figure 2.7 shows a section of the waveform of a musical note.

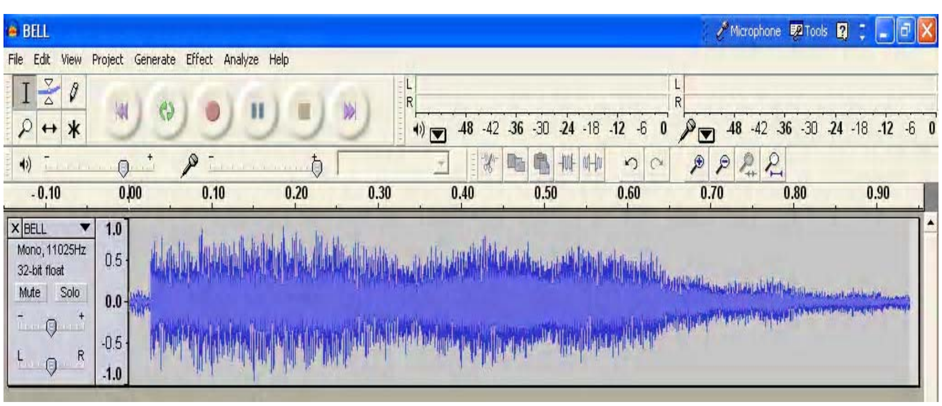


Fig. 2.7 *Waveform View of the Sound of a Bell (from Audacity)*

NOTES

Pure Tone Versus Note

A sound with a single frequency is called a *pure tone*. For instance, if you strike the prong of a tuning fork, the prong of the tuning fork will oscillate at a particular frequency, and it will produce a sound with that particular frequency. The pressure caused by the sound wave can be plotted as continuously changing amplitude on the vertical axis and time on the horizontal axis. If the sound is a pure tone, the graph will be a single-frequency sinusoidal wave. The waveform of a pure tone (440 Hz) sound wave is shown in the Figure 2.8.

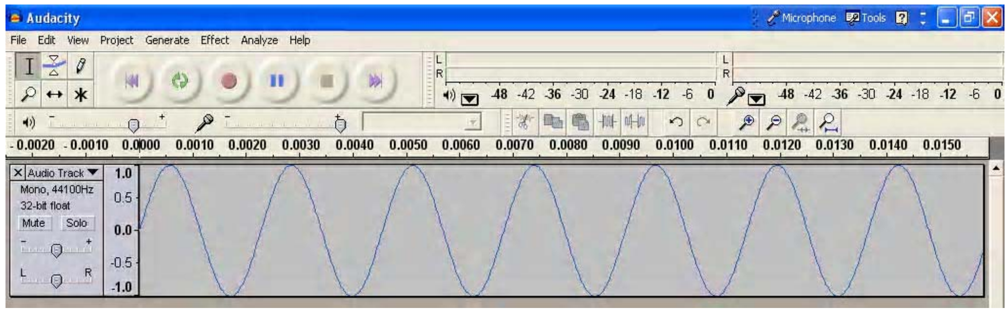


Fig. 2.8 *Waveform View of a Pure Tone at 440 Hz (from Audacity)*

Dynamic Range of Human Hearing

Humans cannot hear sounds of every frequency. The range of hearing for a healthy young person is 20 to 20,000 Hertz (i.e., 20 KHz). This range of frequency is called the *audible range*. This audible range is kept in mind when creating multimedia audio. Also, when digital sound for storage is compressed one often cleverly utilizes this audible range to remove the inaudible frequencies.

The human ear is incredibly sensitive to pressure variations in the air. Atmospheric pressure is generally measured in Newton/Metre^2 (or Pascal), often written as N/M^2 or Pa, and the average atmospheric pressure at sea level is 10^5 Pa. The human ear is capable of detecting pressure variations of the order $2 – 10^{–5}$ Pa (for a pure tone of 1000 Hz), which is less than one billionth of atmospheric pressure. Such a soft sound oscillates the hair cells inside our inner ear with a total peak-to-peak displacement that is much less than the diameter of a hydrogen atom. Our ears are such incredible detectors of sound! For a pure tone of 1KHz, if the pressure amplitude is less than $2 \times 10^{–5}$ Pa, then you cannot hear it. So the *threshold of hearing* can be defined as the sound pressure of $2\times 10^{–5}$ Pa, which is the minimum sound intensity a human with good hearing can detect at 1 KHz in a noiseless environment.

The maximum pressure amplitude the human ear can tolerate is about 28 Pa. The corresponding sound intensity or sound power is about 1 W/m^2. If the sound intensity further increases, it becomes unbearable to the human ear and causes severe pain. The corresponding sound pressure level is known as the *threshold of pain*.

Human Perception of Sound: Pitch, Timbre and Loudness

Pitch is the characteristic of a musical sound by which human ears differentiate between a shrill sound and a dull sound. In other words, pitch is a measure of the frequency of a sound wave. A shrill or high-pitched sound has higher frequency than a flat or dull sound.

Pitch: The characteristic of a musical sound by which human ears differentiate between a shrill sound and a dull sound

Now, a pure tone has a single frequency. However, what about a complex musical note having a range of frequency components? In such case, the pitch is usually close to the fundamental frequency of that sound.

In music, musical notes are ordered from low pitch to high pitch forming a *scale* that provides the basis for a musical composition. The pitch of a note is usually expressed by its fundamental frequency. Musical scales generally consist of seven notes that repeat at the octave.

Timbre is the distinctive quality or tone of a sound that may come from a singing voice or a musical instrument. If you play the same note (say '*Sa*') in flute, harmonium and a grand piano, you will readily identify the instrument even if the sounds are identical in pitch, intensity and duration; provided of course, you have heard the instrument before. It is because our brain can judge the frequency components of the notes and map it with the past experience and make out the identity of the source of sound. The complex manner by which our brain remembers timbre of a note is not fully understood. However, once you have learned to identify a particular timbre, you will be able to identify it even if pitch, intensity and duration are varied. The American National Standards Institute (ANSI) has defined timbre as 'that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar' (ANSI, 1960).

The perception of *loudness* of a sound to the human ear depends mainly upon the pressure amplitude, but it also depends upon duration, frequency, presence or absence of background noises, as well as the sensitivity of the ear. You have learned that the *threshold of hearing* can be defined as the sound pressure of 2×10^{-5} Pa or 0 dBSPL, which is the minimum sound intensity a human with good hearing can detect at 1 KHz in a noiseless environment. If a pure tone at 1 KHz is quieter than this, you will not hear it. To hear a sound with frequency lower than 1 KHz, the sound pressure is required to be more than 2×10^{-5} Pa or 0 dBSPL. For instance, a tone at 50 Hz will not be audible at 2×10^{-5} Pa or 0 dBSPL. In fact such a tone at 50 Hz will be on the threshold of hearing at a sound pressure level of 40 dBspl. Similarly, a 10 KHz tone will be audible at about 5 dBSPL. In other words, a 50 Hz tone at 40 dBSPL, a 1 KHz tone at 0 dBSPL and a 10 kHz tone at about 5 dBspl will have the same perceived level of *equal loudness*.

Masking

Masking occurs when one sound prevents us from hearing a second sound. When you hear a loud sound and a soft sound simultaneously, your ears receive both the sound signals but our brain ignores the soft one and as a result you cannot 'hear' the soft sound. This phenomenon is known as masking. The phenomenon of masking is one of the most important one in psychoacoustics study, since it gives us a basis for eliminating sound information that is not perceived anyway. It helps a great deal in compressing digital audio files.

You are already familiar with the term *threshold of hearing*. It is the minimum sound intensity that a human with good hearing can detect a pure tone *at 1 KHz* in a noiseless environment. This sound level corresponds to 0 dBSPL.

NOTES

Masking: It occurs when one sound prevents us from hearing a second sound

NOTES

This means that the threshold of hearing for a tone at 1 KHz is 0 dBSPL. For other frequencies the threshold of hearing is not 0 dBSPL. The threshold of hearing curve is plotted in Figure 2.9. (a) and (b)

In other words, the threshold of hearing curve gets modified in the vicinity of a loud tone. Watch the following graph carefully. Initially, the sine tone (marked A) of 200 Hz at about 10 dBSPL is below the threshold of hearing curve and hence inaudible. However, the sine tone of 500 Hz (marked B) at about 30 dBSPL is above the threshold of hearing curve and thus, audible.

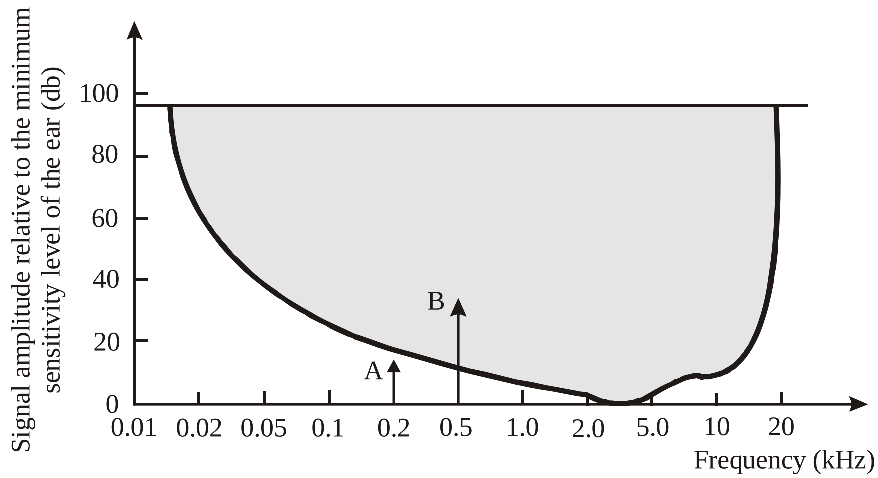


Fig. 2.9 (a) *Threshold of Hearing Curve*

Next, increase the sound level of the 200 Hz tone to 60 dBSPL. The threshold of hearing curve gets modified in the vicinity of the 200 Hz tone and the 500 Hz tone now goes below the modified curve, as a result the 500 Hz tone becomes inaudible (even though it is very much present). The higher intensity sound is called the *masker* and the lower intensity sound is called *masked*. When compressing digital audio signals, the sounds that fall below the threshold of hearing curve may be safely discarded. So the phenomenon of masking can be used cleverly to remove masked data that is inaudible to human ear and thus, reduce file size.

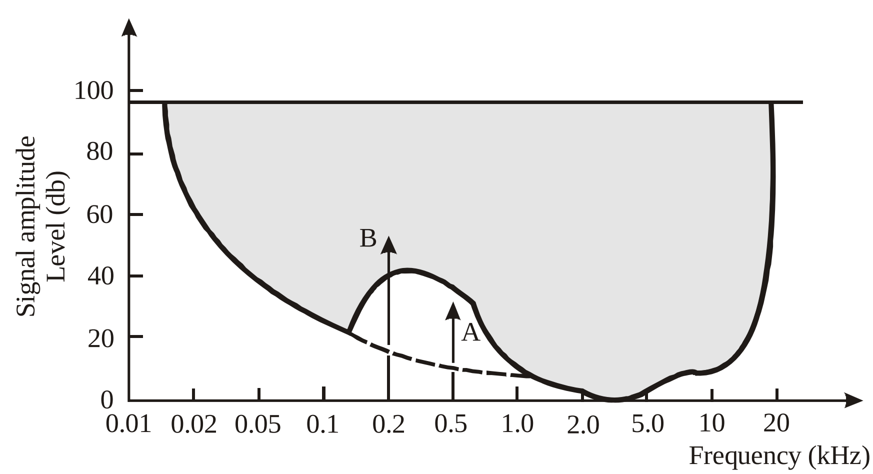


Fig. 2.9 (b) *Threshold of Hearing Curve*

With a particular masker tone you cannot mask all the frequencies. A particular masker can only mask a limited range of frequencies, beyond which masking will not be effective. This range of frequencies is called the *critical band*.

Masking effects are different for different frequencies, narrower for low frequencies, and broader for higher frequencies as is evident from the Figure 2.10.

NOTES

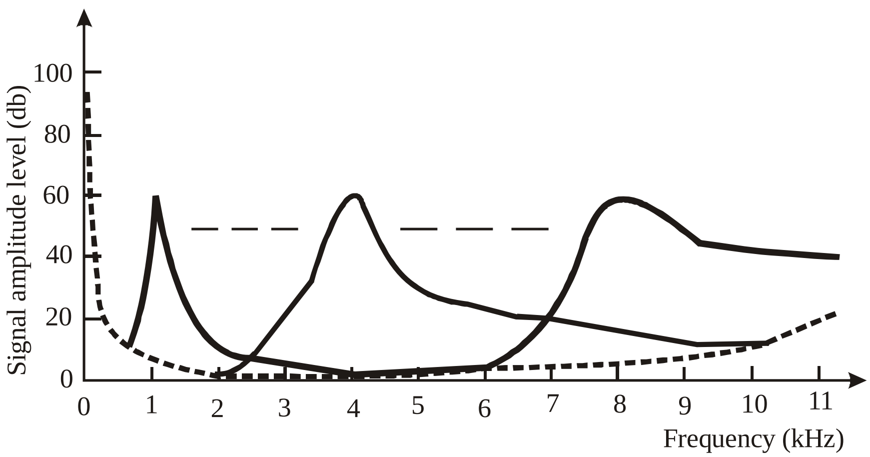


Fig. 2.10 *Effects of Masking*

Elements of Audio Systems

The minimal hardware set-up for digital audio recording consists of the following:
Microphone → Amplifier → A/D Converter → Storage Device (HDD or DAT) → D/A Converter → Speaker

Here, you will learn about the basic features of microphone, amplifiers, sound cards and speakers.

Microphone



Microphone: The microphone is a device used to capture audio waves and convert to electrical signal

The ***microphone*** is a device used to capture audio waves and convert to electrical signal. It is the first component in the preceding chain of components used in a digital audio studio, and in many ways the most important. A microphone is a transducer or a device that converts mechanical energy (travelling waves of compression and rarefaction in a medium, such as, air water, etc.,) into patterns of electrical current. Despite there being many different categories, you can divide microphones into two distinct classes based on their working principles: dynamic microphones and condenser microphones. Each of these classes contains numerous different designs and several variations of those designs.

Dynamic or moving coil microphones work by vibrating a thin metallic diaphragm and an attached coil of wire in a magnetic field. Generally, the thin diaphragm is attached to a coil of wire that surrounds or is surrounded by a high-powered magnet. As a magnet goes through a wire or a coil of wire, the magnet induces current to flow in the wire. Sound wave causes the diaphragm and the coil to move in the magnetic field that in turn generates small amount of electrical current or voltage. Thus, dynamic microphones work by electromagnetic induction. More the intensity of the sound, more is the electric current induced in the coil. Moving coil microphones are very popular. They are uncomplicated, cheap and robust. As a result, they are most often

used for field works. They are suitable for both voice and loud instruments, such as drums, etc.

NOTES

Another type of dynamic microphones is *ribbon type microphone*. Here a very light ribbon of metal, usually corrugated in shape, is suspended in a powerful magnetic field. The vibration of the ribbon within the magnetic field, as sound wave oscillates it, generates electrical signal due to electromagnetic induction. Ribbon microphones give very high quality sound reproduction but are usually very fragile due to their inherent construction. However, in a studio they can be very effective for both voice and most instruments.

Condenser microphones work on the principle of electrostatic induction. If two oppositely charged plates are moved closer or away from each other, they will result in a flow of current. In the condenser microphone, the very thin diaphragm plate usually coated with metal, functions as one of the plates of a capacitor. The movement of the diaphragm plates due to incident sound wave changes the capacitance of the capacitor. A rigid back plate serves as the other plate of the capacitor. An electrical potential or charge is maintained between these two plates by means of an external DC source (phantom power) or by a battery. As the distance between the plates changes due to the attack of the sound wave, current flows in the wire. The amount of current is proportional to the intensity of the sound or in other words, the displacement of the diaphragm. Unlike the moving coil microphone, the thin and lightweight diaphragm of the condenser (without an attached metal coil) allows the condenser microphone to respond better to fast transient sounds and to high frequencies. Condenser microphones are becoming increasingly popular for general recording.

Audio Amplifier

Recall the following chain of components that form minimal hardware set-up for digital audio recording and playback:

Microphone → **Amplifier** → A/D Converter → Storage Device (HDD or DAT) → D/A Converter → Speaker

An *audio amplifier* is a vital component in the preceding audio recording and playback chain. The audio amplifier takes as input low-power audio signals, corresponding to sound frequency within the human range of hearing (between 20 Hz to 20,000 Hz) and amplifies it to a level suitable for driving speakers. An audio amplifier is an electronic device that uses a series of transistors incorporated into integrated circuit chips as their primary component.

Loudspeaker

Loudspeaker: The last component in the preceding audio playback chain

Microphone → Amplifier → A/D Converter à Storage Device (HDD or DAT) → D/A Converter → **Loudspeaker**

A ***loudspeaker*** is the last component in the preceding audio playback chain. The loudspeaker is an analog device that converts electrical energy back to sound energy. The digital to analog converter (D/A converter) sends fluctuating electrical current to a coil attached to a flexible *cone* or *diaphragm* (made of paper, plastic or metal) and placed in a magnetic field. As the current flows in the coil, attraction and repulsion between the magnetic fields created by the fluctuating current and the

permanent magnet produces vibration of the flexible cone or diaphragm. This vibrates the air in front of the cone or the diaphragm, creating sound waves. Thus, the electrical audio signal is converted to sound wave.

NOTES

Converters

Analog-to-Digital Converter (ADC): The ADC translates the analog sound waves into digital data that the computer can recognize. The analog audio signal received from the microphone or the line-in port is sampled, quantized and the code word is generated in binary form by taking precise measurements of the wave at frequent intervals. Sound cards contain at least one ADC converter for each of the stereo channels. Often the sound card contains more than two ADCs.

Digital-to-Analog Converter (DAC): The DAC converts recorded or generated digital audio signal to analog form for playback. Sound cards contain at least one DAC converter for each of the stereo channels (just as they do for the ADCs). Often the sound card contains more than two DACs.

Some sound cards, instead of separate ADCs and DACs, use a COder/DECoder chip (CODEC) that performs both the preceding functions.

Even if you can sample audio signals at as high a rate as 96 Khz or 192 Khz, it is often impractical to sample the audio at that high rate, as sampling at 44.1 Khz effectively covers all the frequencies within the human audible range, and moreover, the file sizes become very large.

Digital Signal Processor

The *Digital Signal Processor* (DSP) is a special type of microprocessor chip designed to manipulate or process digital signals (audio, image, video, etc.), which have been converted from analog form by the ADC. In a sound card, the DSP used is dedicated to take the load off the CPU of the computer and performs the mathematical processing involved in digital audio signal processing.

The DSP determines how many Musical Instrument Digital Interface (MIDI) channels, sound streams or voices the sound card can support. Hence, it is a key component of the sound card. In the world of synthesizers and audio signal processing, the term *voice* is used to indicate a single note from a single instrument. For example, if a sound card is simultaneously playing back a note from a saxophone and two notes from piano, then the sound card is reproducing one plus two, i.e., three (3) *voices* and as a result three notes are played at the same time. Modern sound cards can simultaneously handle 64 *voices* (also known as 64 *voice* polyphony).

Memory Bank

This is the local memory of the sound card for storing audio data during digitization and play back of sound files. Single In-line Memory Module (SIMM) memory banks were used in earlier sound cards, whereas the new ones use Dual In-line Memory Modules (DIMM).

Connectors

A sound card typically provides the following external ports:

NOTES

Line-Out: *Line-out* is an un-amplified stereo output that you can connect to amplified stereo speakers, headphones, tape recorder or DAT recorder, etc. Conventionally, the line-out port is given lime color. A sound card may have multiple line-out ports. For example, sound cards that support four speakers may have two stereo line-out ports.

Microphone-In (or Mic)**:** The microphone-in port (commonly referred as ***mic***) is used for connecting a personal computer microphone for voice or musical input. Conventionally, the ***mic*** is given pink color.

Line-In**:** *Line-in* is an un-amplified stereo input used to receive audio source from a record/playback device, such as CD-player or a VCR. The standard color for line-in is light blue.

MIDI/Gameport**:** The *MIDI/gameport* connector is used to connect external MIDI devices, such as synthesizers and keyboards and also game controller using a special type of cable. The standard color is gold.

Speaker-Out/Subwoofer**:** The speaker/subwoofer connector is used to connect un-powered speakers or in some cases to powered subwoofers that require a high-level audio signal input. The standard color is orange.

Wavetable/FM Synthesizer

This chip processes the MIDI instructions sent by the synthesizer/keyboard for generating sound. You will learn about this in detail later on in this unit. The MIDI wave-table chip utilizes bits of prerecorded digital sounds. On the other hand, the FM synthesizer chip, which has almost become obsolete, now generates the sound by combining pure tones.

Amplifier

Most high-end professional sound cards do not have the built-in amplifiers, because they cater to amplified speakers. However, in the early years of sound cards, mostly un-amplified speakers were used, and the audio signal had to be amplified in the sound card itself before they were sent to the speaker for play back. The legacy continues and you will still find an amplifier circuit built in the low or mid-range sound cards to cater to playback through the un-amplified speakers.

Musical Instrument Digital Interface

You have seen how analog sound is sampled and quantized to convert it to digital audio form. Here you will study about yet another way to store information related to sound in digital form. It is called Musical Instrument Digital Interface (MIDI).

Unlike, a digital audio file (for example, a *.wav* or a *.snd* file) a MIDI file contains information regarding the instrument, the note to be played, the duration of play, etc. In other words, written in a scripting language, a MIDI file contains the details of each *event* or keystroke, the change of note, tempo, etc., as a musician plays a synthesizer emulating a flute, a bass, a piano or a saxophone.

When contrasted with sampled audio files MIDI data files sizes are very small. For example, a digital audio file created at 44.1 Khz with 16-bit resolution and two (stereo) channels (require by files containing high-quality stereo sampled audio) generates about 10.3 MB of data per minute of sound, while a typical MIDI sequence might generate even less than 10 Kbytes of data per minute of sound as it consists of

only the commands needed by a synthesizer to play the sounds and not the sampled audio data. These commands are in the type of MIDI messages that instruct the synthesizer which sound to use, which notes to play, and how loud to play each note, etc. Then the synthesizer produces the actual sounds.

NOTES

As MIDI is merely a script describing the musical composition. One advantage of using MIDI is the ability to easily edit the script and if required change the playback speed or the pitch or key of the sound independently. So depending on the requirement, the tempo of a musical composition may be increased or decreased easily to suit that of the singer.

These days MIDI instruments with sophisticated wave-table synthesis facilities have revolutionized the digital sound recording studios and their functioning.

MIDI versus Sampled Audio

You are aware that, written in a scripting language, a MIDI file contains information regarding the notes and duration of notes along with the instrument that will be emulated to play the note. Each such piece of information or message is called an *event* in MIDI terminology.

On the other hand, a sampled digital audio contains large number of quantized samples (for example, 44,100 samples per second for a mono sound track) so that when played back they can reproduce the waveform of the original sound. So, inherently the sampled digital audio files are much larger in size than MIDI files playing the same music for identical duration.

Also, as MIDI files stores information in terms of the instruments played and the details of the notes, editing of MIDI music is much easier. You may even edit the MIDI script such that a particular portion (or even the whole piece of music) of the music is played in flute instead of say a piano. This is not exactly possible with sampled digital audio, as there is no discrete start and termination of each musical note in the audio file.

However, sampled digital audio files, those that are sampled and quantized properly (for example, 44.1Khz /16bit depth for CD-quality audio), will capture and play back sound just as it is played with delicate characteristics, such as timbre, changes in pitch, tone, etc. In other words, the digitally recorded audio retains the characteristics of the instrument and the musician. As MIDI audio uses synthesized sounds to recreate each note played on the flute, the same MIDI note will play back exactly the same sound again and again. To well-trained ears, a MIDI sound may sound artificial and synthetic than sampled digital audio.

This drawback of MIDI has been corrected to some extent by including additional information that can be included in the MIDI event, such as how a note should be modulated or bent. The events also contain information as to how hard a note is played (how gently or firmly a piano key is pressed to play a particular note), so the subtle individuality of a master musician is preserved.

If you know how to play a MIDI keyboard or a synthesizer, it will be much easier for you to create (and edit) a MIDI audio than sampled audio. Simply connect a MIDI keyboard to the computer, play a musical piece, record the music in MIDI format, and edit it with any editing program that can edit MIDI. Suppose, you have

NOTES

created the musical piece using violin and suddenly feel that the use of flute would have been much better; just select the piece and change the instrument type from violin to flute at the click of the mouse. You can also edit individual notes, change the key in which the piece is played, and fix errors note by note. These are distinct advantages of MIDI over sampled audio.

Also, as MIDI format is now an industry standard, MIDI music can be easily transported from one MIDI device or computer to another.

MIDI Standards

MIDI is a standardized protocol or procedure set. Manufacturers of musical instruments, computers and computer software now routinely adopt MIDI protocol. MIDI information is transmitted as 'MIDI messages' or instructions telling a music synthesizer how to play a piece of music. The MIDI protocol defines how these MIDI messages are constructed, transmitted, stored, and what they denote. As MIDI message is the key building block for MIDI music, the protocol or standard defining the MIDI messages have been firmly established. The hardware part of the MIDI protocol stipulates how two MIDI devices are connected, how MIDI ports convert data to electrical signals and how the MIDI cables transmit the signal. The software part of the MIDI protocol specifies the format and specification of MIDI messages.

MIDI Hardware and Software

The following are the various MIDI Hardware and Software:

MIDI Controllers

You have learned that to generate MIDI sound, you have to generate MIDI messages. The hardware devices that generate MIDI messages are called *MIDI controllers*. MIDI controllers can be of various types. Digital musical instruments, such as a keyboard or a guitar can serve as a MIDI controller if it is designed for generating MIDI messages.

A MIDI keyboard resembles a piano keyboard and have some extra controls. The number of keys, controls and sensitivity features varies from keyboard to keyboard. A standard keyboard has 88 keys. The MIDI keyboards often come with features, such as *touch sensitivity* or detection of velocity of the keystroke, *after-touch* or how hard you hold down a key and embed this information to the MIDI message.



Fig. 2.11 *A MIDI Keyboard*

MIDI Synthesizers

On the other hand, devices that can read MIDI messages and convert them into audio signals for play back through an output device are called *MIDI synthesizers*.

NOTES

Some MIDI keyboards can serve as both synthesizers and controllers, that is, they can both play back sound by interpreting MIDI messages as well as generate MIDI messages.

Some MIDI keyboards are capable of generating MIDI messages. They are not equipped to produce any sound. You can select such a silent MIDI keyboard and connect it with your computer either having a sound card that is equipped to synthesize MIDI audio or having a MIDI software synthesizer provided by the operating system.

MIDI Cabling

Depending on the connection type (port) available with the computer, different types of MIDI cables are available. Earlier computers used a 15-pin MIDI/joystick connection. These days MIDI cables are used connect to the USB port of the computer. The MIDI-device side of the MIDI cable uses two 5-pin DIN connectors, one for the *in* and one for the *out* port. A through port is also available on some MIDI devices to pass-over MIDI messages directly through to another MIDI device. A standard MIDI connection passes data serially at a rate of 31.25 Kbits/sec. A computer can address multiple MIDI devices at the same time using high-speed serial ports.

MIDI Sequencer

Earlier a hardware device called *MIDI sequencer* was used to receive store and edit MIDI data. However, now software application programs are available that can perform the same tasks using a personal computer. You will assume that you will be using a software sequencer program to edit and store MIDI data. Cubase or Cakewalk are examples of such MIDI sequencer programs. The sequencer program receives the MIDI messages generated by the MIDI controller (for example, MIDI keyboard) and stores them in General MIDI file format (.mid). Many MIDI sequencer programs allow viewing the MIDI file in different formats including musical notation or an event list for easy editing purpose.

MIDI Channel and MIDI Track

A MIDI *channel* is a path for data communication between two MIDI devices. The solitary physical MIDI channel is split into 16 logical channels by allocating a 4-bit channel number inside the MIDI messages. A MIDI controller keyboard can be set to transmit using any one of the 16 MIDI channels. Hence, the channel is the path along which the MIDI messages are passed from the computer to the keyboard or some other playback device, such as the synthesizer, etc.

A *track*, on the other hand, is an area in memory where the MIDI data is stored. In a MIDI sequencer, you can use the track view-to-view and edit each track separately. Recording has to be done on a specific track. While using multiple tracks, individual tracks can be listened or muted separately.

The concept of channels and tracks should be clear. The General MIDI (GM) standard stipulates 16 channels on a MIDI device. However, in your MIDI sequencing software you are to decide which track you will designate to which channel. For example, you can designate track 1 to play on channel 11 and track 2 to play channel

NOTES

1. It is up to you but assign them different channels. If, otherwise, you assign both the tracks the same channel number you will hear only one of the two instruments playing in the two tracks.

As per the GM standard, channel 10 is designated to carry drum and percussion sounds. Hence, if you set a track to record from channel 10, then each key you play at the keyboard will be recorded as some drum or percussion instrument.

MIDI File Formats

The MIDI file format contains MIDI messages as discussed earlier. The MIDI message is contained in two chunks of data, header chunks and track chunks**.** These MIDI messages can be collected and stored in a computer file system, in what is commonly called a standard MIDI file (SMF) or more commonly called a MIDI file. The SMF specification was developed and maintained by the MIDI Manufacturers Association (MMA).

Huge collection of SMFs are available on the web, most commonly with the extension ***.mid***. MIDI-Karaoke (which uses the ***.kar*** file extension) files are an unofficial extension of MIDI files and are used to add synchronized lyrics to standard MIDI files.

Audio File Formats

You have seen that digital audio files are saved in wide range of formats having different file extensions, such as .wav, .mp3, .au, .rm, and so on. The right choice of audio file type for any specific multimedia application is very important. To select the right audio file type you should be able to identify the file types and differentiate them.

Common Audio Formats

The following are some common audio formats:

- *WAV (.wav)***:** This audio format is chosen as the native format by Microsoft® for all Windows operating systems. Almost every browser has built-in WAV playback support and a number of CODECs supporting .wav files.
- *MP3 (.mp3)***:** As already discussed, *mp3* is the name of the file extension and also the name of the file type for MPEG-1, Audio Layer-III. The Layer-III coding scheme employs perceptual audio coding and psychoacoustics compression to eliminate all unnecessary sound that the human ear cannot hear without sacrificing sound quality. The mp3 CODEC is a copyrighted one and cannot be used to compress digital audio without licence.
- *Windows Media Audio (.wma)***:** It is a Microsoft® file format for encoding digital audio files akin to MP3. It can condense files at a higher rate than MP3. Since the WMA files uses the *.wma* file extension, they can be compressed to go with diverse connection speeds or bandwidths.
- *Real Audio (.ra .ram .rm)***:** Real audio is a proprietary format developed by Real Networks Inc. It is used for streaming audio that enables you to play digital audio files in real-time on the Web. However, to use this type of file you must have Real Player installed in the PC which is, however, freely downloadable.
- *MIDI (.mid)***:** You have already studied the MIDI files in details. The file extension is ***.mid***.

- *Audio Interchange File Format (AIFF):* Audio Interchange File Format (AIFF) is an audio file format standard used to facilitate file exchange. The AIFF files are divided into chunks, each with its own header and data. The *.wav* and *.aif* (Apple) are variants of the AIFF format.

NOTES

2.3.2 Video

In the previous sections, you have learned how text, image or audio can be used in multimedia. In this section, you will learn how video works as a key multimedia object along with the different formats and standards of multimedia video. Digital video is perhaps the most prominent multimedia object to create impact on its audience. With the advancement of cheaper storage technology and better video compression techniques, digital video has become a very widely used multimedia object. The term '*video*' has been derived from Latin, meaning 'I see' or 'I apprehend'. The term videography refers to the process of capturing moving pictures.

Video technology has evolved from television technology, but it has now developed to a great extent to allow consumer digital video recording and playback. The standards that have evolved initially through analog television and film, then digital video, and finally to present day HDTV have jumbled up with lot of ambiguous terminologies and nomenclature. To get rid of this ambiguity, you have to know the basics of analog video, television technology and film. These are the three close kins of digital video and have many features in common. Moreover, to understand the mechanism of digital data communication you have to understand how data is communicated in an analog manner. You have to understand the concepts of frame, number of lines in a frame, frame rate, etc. These concepts evolved from analog television. Knowing them makes it easier to understand the similar issues in digital video. Another similarity between television and digital video technology is their requirement for bandwidth. Just as the television in the analog domain requires bandwidth in the airwaves, the digital video requires bandwidth for transmission across the computer network. Bandwidth is a costly component, so its requirement is to be controlled and minimized to keep things economical. You will see what factors influences the bandwidth and how to control them for digital video. Digital technology is all around us, so you should have a fair idea about the frequently used terminologies, such as television screen sizes, resolution, etc.

Video Camera

To understand how digital video camera functions, you should first study some fundamental aspects of analog video camera.

Analog Video Camera

In an analog video camera, light comes through the lens and hits an imaging chip, which reacts to the light with continuously varying voltages. The stronger the light, the stronger the voltage is. These voltages, after magnification and signal processing, magnetize the tape particles in a continuously varying (analog) pattern that stores the signal.

In analog systems, the video signal from the camera is delivered to the video through the connector cables of a VCR, where it is recorded on magnetic videotape.

NOTES

The video signal is written to tape by a spinning recording head that changes the local magnetic properties of the tape's surface in a series of long diagonal stripes. The head is tilted at a slight angle compared with the path of the tape; it follows a helical (spiral) path and which is called 'helical scan' recording. Each stripe represents information for one field of a video frame.

You know that color models divide color into three components. In case of digital still images, you can use the RGB color model—where a color is separated into its red, green, and blue components. In the analog video, luminance/chrominance models, such as the YUV and YIQ, it functions better. The luminance — chrominance color models also need three pieces of information — one luminance (value or brightness) and two chrominances to represent a color. In case of an analog video, the information can be passed on in either of three ways — in *component*, *s-video* or *composite* form.

In *component video*, separate signals are sent for each part of the three luminance-chrominance components. Component video has three separate paths for the information and three connectors at the end. This is the most accurate format for representing and transmitting video, as crosstalk between the different components is eliminated. However, it is also very expensive format. Only a few high-end analog video cameras like Betacam use component video connections (see Figure 2.12).



Fig. 2.12 *Typical RGB Component Video* Connection

The next way to transmit video is *s-video*, which utilizes two data paths, one for the luminance component, and the other for two chrominance components.

In *composite video* the video signal is sent on just one channel. The signal is sent on a single channel by compositing the signal. The main disadvantage to this technology is that the quality of the signal may deteriorate due to crosstalk between the color (chrominance) and luminance components. Thus, composite video is the lowest quality of all the three alternatives.

Some of the popular type of analog video camera formats during the 1980s was VHS, S-VHS and Hi-8. S-VHS had much higher resolution and higher color quality support than VHS. The Hi-8 video uses a smaller tape, and thus video cameras were smaller than for VHS. Then came the *Betacams*. The Betacams used small size

tapes and gave extremely high quality video and became very popular in the news reporting circuit.

NOTES

The resolution of any given video camera depends on whether it is NTSC-, PAL- or SECAM compliant. The number of horizontal lines in a frame is called the *vertical resolution* of an image. Now, for an analog video camera, the picture information is sent as a continuous waveform rather than as discrete pixels in case of the digital video camera. So the horizontal resolution for an analog video camera is only indicative. It is not fixed but lies within a range. That is why you will find that the preceding standards specify only the vertical resolution, but no horizontal resolution. The analog video formats are summarized in the Table 2.2.

Table 2.2 *Analog Video Formats*

Video Format	**Year Introduced**	**Colour Transmission Format**	**Horizontal Resolution**	**Tape Width**	**Quality**
VHS	1976	composite	~240	½" (12.5 mm)	consumer
Betamax	1976	composite	~240	½" (12.5 mm)	consumer
8mm (Video 8)	1984	composite	~240?300	8 mm	consumer
S-VHS	1987	s-video	~400?425	½" (12.5 mm)	high-end
consumer					
Hi-8	1998	s-video	~400?425	8 mm	high-end
consumer					
U-Matic	1971	composite	~250?340	¾" (18.75 mm)	professional
M-II	1986	component	~400?440	½" (12.5 mm)	professional
Betacam	1982	component	~300?320	½" (12.5 mm)	consumer
Betacam SP	1986	component	~340?360	½" (12.5 mm)	professional

Note: Vertical resolution depends on whether the camera is NTSC-, PAL- or SECAM-compliant.

Analog video has become outdated with the availability of digital video cameras at affordable price for personal and professional use. However, analog video cameras can still be connected to the PC with a video capture card and appropriate software—and video clippings can be captured in digital format directly on the computer. Also, using an interface device and a VHS player, an old VHS tape can be played and captured digitally on the computer. Video capturing is the processing of encoding and saving video in digital format on a computer.

Digital Video Camera

Just like the analog video camera, digital video cameras detects light coming in through the lens and scans the image line-by-line. However, here the light intensity is digitally encoded, unlike the analog camera, where the light is recorded as continuously changing voltage. Just like the analog camera, the encoded pixel values can also be stored on videotape. However, the information is analog and thus different from the analog signal.

NOTES

A unique sensor converts light into an electronic signal, which is called Charge-Coupled-Device (CCD), when the light reflected from a subject passes through a video camera lens. CCD was invented way back in 1969 at Bell Labs, by George Smith and Willard Boyle and is used in many digital optical devices including camcorders and digital still cameras to convert light energy to electronic signal. They are also used in astronomical telescopes, scanners and bar code readers. CCD is a light-sensitive integrated circuit. It is at times referred to as a *chip* or microchip that stores and shows the data for an image in such a way that each pixel (picture element) in the image is changed into an electrical charge, the intensity of which is related to a color in the color spectrum. The camera processes the output of the CCD into a signal having three channels of color information and synchronization pulses (sync). Different video standards are there for managing the CCD output, each deal with the amount of separation between the components of the signal.

Some high-quality cameras used for professional broadcasting have as many as three CCDs (one for each color of red, green and blue) to augment the resolution of the camera (see Figure 2.13).

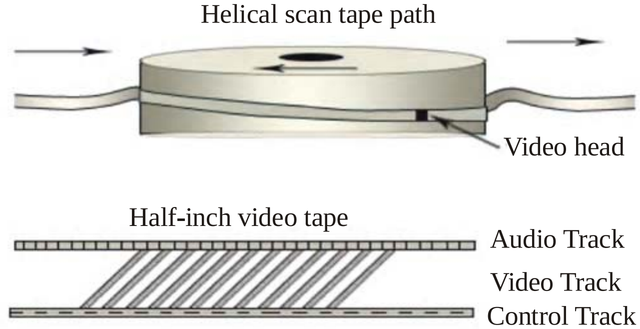


Fig. 2.13 *CCDs for High Quality Cameras*

Transmission of Video Signals

Basically, video or motion pictures are created by displaying images depicting progressive stages of motion at a rate fast enough so that the projection of individual images overlap on the eye. *Persistence of vision* of human eye, which allows any projected image to persist for 40-50 ms, requires a frame rate of 25-30 frames per second to ensure perception of smooth motion picture.

In a video display:

- ***Horizontal resolution*** is the number of distinct vertical lines that can be produced in a frame.
- *Vertical resolution* is the number of horizontal scan lines in a frame.
- *Aspect ratio* is the width-to-height ratio of a frame.
- *Interlace*[†] *ratio* is the ratio of the frame rate to the field rate.



Horizontal resolution: The number of distinct vertical lines that can be produced in a frame

Constitution-wise there are three types of video signals—*component video*, *composite video* and s-*video*. Most computer systems and high-end video systems

[†] In Interlacing, each frame is divided into two fields–odd and even, each consisting of alternate horizontal lines. Every frame is refreshed fully by refreshing the two fields alternately. Thus, flicker-free image is displayed at a low refresh rate or bandwidth.

use component video whereby the three signals R, G and B are transmitted through three separate wires corresponding to red, green and blue image planes, respectively. However, because of the complexities of transmitting the three signals of component video in exact synchronism and relationship these signals are encoded using a frequency-interleaving scheme into a composite format that can be transmitted through a single cable. Such format known as composite video, used by most video systems and broadcast TV, uses one luminance and two chrominance signals. *Luminance* (Y) is a monochrome video signal that controls only the brightness of an image. *Chrominance* is actually two signals (I and Q or U and V), called color differences (B–Y, R–Y) and contains color information of an image. Each chrominance component is allocated half as much bandwidth as the luminance, a form of analog data compression*, which is justified by the fact that human eyes are less sensitive to variations in color than to variations in brightness. Theoretically, there are infinity of possible combinations (additive) of R, G and B signals to produce Y, I and Q or Y, U and V signals. The common CCIR 601 standard defines:

NOTES

Luminance (Y) = 0.299R + 0.587 G + 0.114B

Chrominance (U) = 0.596R – 0.247 G – 0.322B

Chrominance (V) = 0.211R – 0.523G + 0.312B

The inverse of the preceding transformation formula gives

Red (R) = 1.0 Y + 0.956 U +0.621 V

Green (G) = 1.0 Y - 0.272 U -0.647 V

Blue (B) = 1.0 Y - 1.061 U -1.703 V

Unlike composite video, s-video (separated video or super video as S–VHS) uses two wires, one for luminance and another for a composite chrominance signal. Component video gives the best output since there is no crosstalk or interference between the different channels unlike composite video or s-video.

Digital Video Standards

To improve picture quality and transmission efficiency, new generation televisions systems are designed based on international standards that exploit the advantage of digital signal processing. These standards include *High Definition Television* or *HDTV*, *Improved Definition Television* or *IDTV*, *Double Multiplexed Analog Components* or *D2-MAC*, *Advanced Compatible Television, First System* or *ACTV-I*. The HDTV standard that support progressive (non-interlaced) video scanning has much wider aspect ratio (16:9 instead of 4:3), greater field of view, higher horizontal and vertical resolution (9600 and 675 respectively in the USA) and more bandwidth (9 MHz in USA) as compared to conventional color TV systems.

Video File Formats

In a television transmission system, every part of every moving image is converted into analog electronic signals and transmitted. The VCR can store TV signals on magnetic tapes, which can be played to reproduce stored images. There are three main standards

* Known as *chroma subsampling* where Y's resolution is 4 times than U's and than V's resolution (horizontal 2 and vertical 2).

NOTES

for analog video signals used in television transmission: NTSC, SECAM and PAL. Television standard in India is based on the Phase Alternate Lines (PAL) system. The PAL system is followed in the countries with Alternating Current (AC) frequency of 50Hz, such as in UK, Western Europe, Australia, China, India, South Africa and South America. In this system, the screen resolution is 625 lines but the scan rate is 25 frames per second (to suit the AC frequency). The National Television Standards Committee guided the development of America's television standard hence it is called NTSC. Another 50Hz Standard is Sequential Color Avec Memoire (SECAM). This standard is mainly followed in France, the Eastern Block and Middle Eastern Countries.

2.3.3 Animation

Literally speaking, to *animate* is to bring to life, i.e., to put something into action. Animation makes graphics more realistic by imparting motion and dimension to an inanimate object. Intuitively though we think of animation synonymous with motion, technically speaking, it covers all changes that have a visual effect. Thus it may include time varying position (motion dynamics), shape, size, color, texture (update dynamics) of an object and also changes in lighting, camera position, focus, etc.

Application

With advancement in computer-aided techniques, today animation is extensively used in Entertainment (games and movies), Educational and Training presentations, Advertising, Internet and Process simulation. Process simulation through animation is very useful in visualization of functioning and stages of operations of industrial products (like a gear or motor) or, gradual transformations in a complex process (like changing atomic structures in a chemical reaction, or, distortion of structures under dynamic forces).

Elements

A computer animation sequence can be set up by specifying the storyboard, the object definitions, and the image frames. The ***storyboard*** is an outline of action. It could consist of rough sketches of motion sequence or it could be a list of basic events that are to take place. ***Object definitions*** are given for each participating object in terms of their shape and movement. The still image frames are either drawn manually or computer-generated to simulate motion-sequence of animating objects. The illusion of movement is created by playing 15-20 numbers of such still images with small changes made to each one per second. The eyes retain an image long enough to allow the brain to connect the frames in an uninterrupted sequence. In traditional animation, as many as 30 FPS might be used to give a smoother appearance at high speeds.

Elements: A computer animation sequence can be set up by specifying the storyboard, the object definitions, and the image frames

Animation Techniques

The following are the various animation techniques:

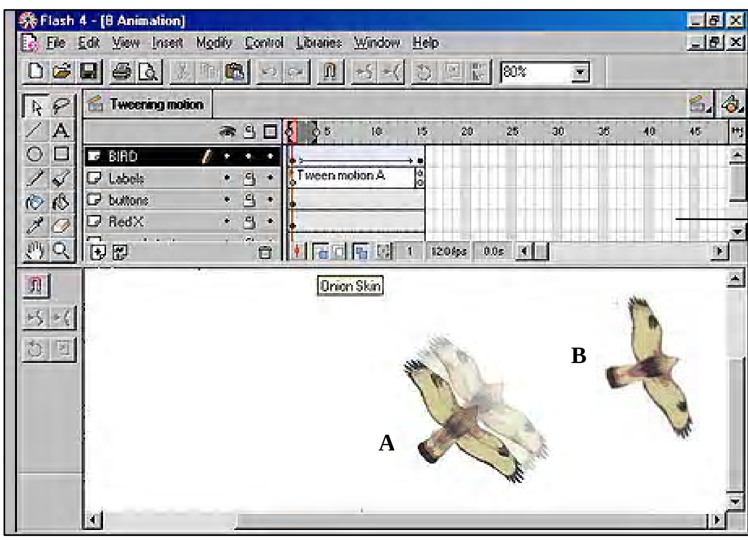
Cel Animation

Classically, picture frames depicting animated sequence were drawn manually. In doing so the ***onionskin*** technique which is popularly known as c*el animation* technique is mostly adopted. By drawing on a onionskin-like translucent paper called 'cel', with a

light source beneath the drawing surface, an animator can see the position of an object on one page, while drawing it in a new position on the page above. Only the moving elements on the cel need to be redrawn for each frame while the fixed part (usually the background) of the scene need only be made once.

NOTES

This concept of cel has been implemented in the digital media in the form of ***layer***. Many animation softwares offer translucent drawing layers that are shown progressively more opaque, to assist in identifying the stacking order of the layers. The image frames of an animated sequence can be made by combining a background layer, which remains static, with one or more animation layers, in which any changes that take place between frames are made. To take a simple example, suppose we wish to animate the flying of a bind in the sky. The first frame could consist of a background layer containing the and a foreground layer with an image of the sky. To create the next frame, we would copy these two layers and then, using the move tool, displace only the bird's image a small amount. By continuing in this way, we could produce a sequence depicting the smooth movement of the bird across the background (see Figure 2.14).



Here you see the Macromedia Flash time line. You can see layers of objects stacked up for display as a single image. The numbered columns each represent a frame, with time marching on from left to right.

In this animation, as the bird flies from A to B, two intermediate frames are displayed with onion skin effect. Thus all the in-betweens can be displayed in onion skin mode and can be edited on frame by frame basis if required.

Fig. 2.14 *Flying of a Bird using Cel Animation*

The form of animation based on objects movement only (no change of other properties) is called ***sprite animation***. The moving objects are referred to as *sprites*. Instead of storing changes of sprite position from frame to frame the change values can also be generated by computer programs.

Keyframe Animation

Keyframes are image frames that depict the key positions of the objects being animated and marks significant changes in the animation sequence. Usually the extremes of an action or sequence like start, stop and changes of movement direction occur at keyframes. The more intricate and rapidly varying the motions are, the more number of keyframes are required. ***In-betweens*** are the intermediate frames drawn between the keyframes and are used to smooth the transition from one keyframe to the next.

Keyframes: Image frames that depict the key positions of the objects being animated and marks significant changes in the animation sequence

NOTES

We are familiar with traditional cartoon animation used in the movies or televisions where artists meticulously draw each frame of a scene and then capture them frame by frame with a movie camera. While the expert artist draws the keyframes, the assistants create the filler in-betweens most mechanically.

In a computer-based animation, standard software tools are available to design the keyframes. The software then figures out the in-betweens by applying *interpolation* algorithms. One does not have to design every intermediate frame individually. The process of generating the in-betweens is commonly known as ***tweening***. The most basic interpolation technique is ***linear interpolation* or *lerp***. Given the values p_1 and p_2 of some attribute (position, color, size) in two keyframes corresponding to times t_1 and t_2 respectively, the value p_t at any intermediate frame corresponding to time t is given by

$$p_t = \frac{t_2 - 1}{t_2 - t_1} p_1 + \frac{1 - t_1}{t_2 - t_1} p_2.$$

This implies $p_t = (1 - n)\, p_1 + n\, p_2;$

where $n = \dfrac{1 - t_1}{t_2 - t_1} \leq 1 \text{ and } \geq 0.$

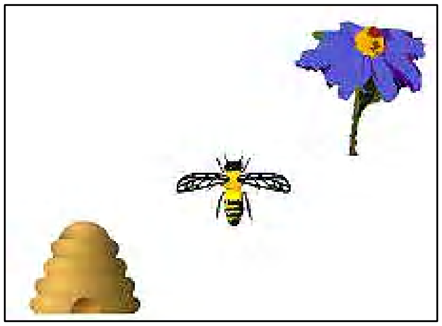
Lerping generates motion that starts and stops instantaneously, with objects attaining their full velocity as soon as they start to move, and maintaining it until they stop. For such constant velocity animation equal-interval time spacing is used. Moreover the motion path is linear which cannot simulate the realistic curvilinear trajectory (for example, projectile path) of moving objects.

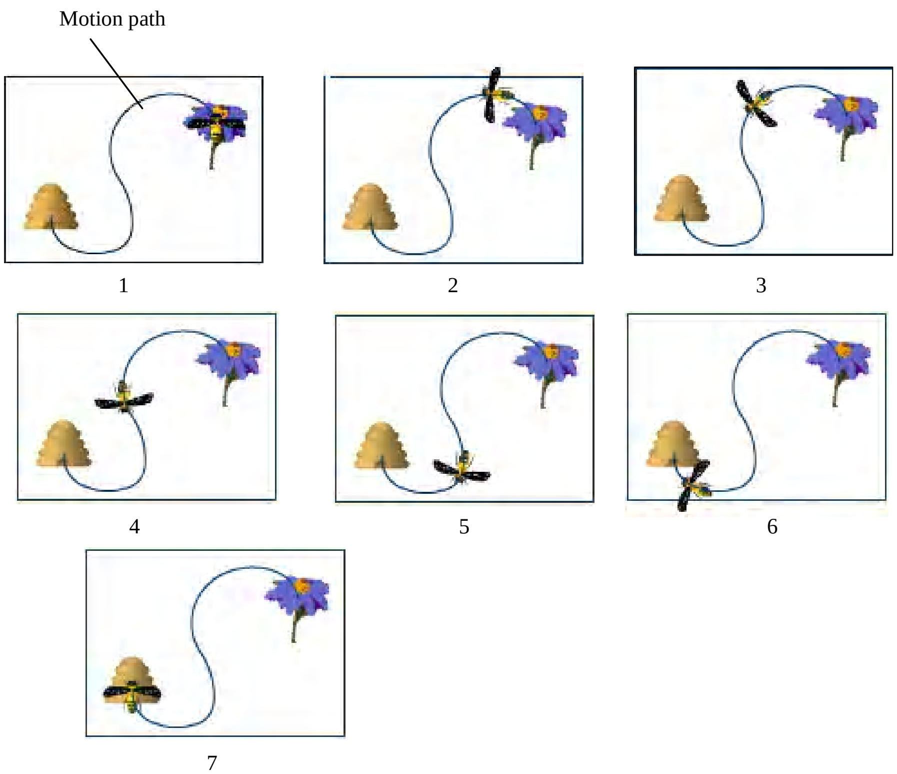
To make different parameters vary realistically with time ***spline interpolation*** technique is often used. By using Bezier functions instead of linear function to interpolate between keyframes, smooth motion can be achieved simulating the gradual increase of velocity at the start and conversely gradual decrease of velocity at the end. When a motion begins, the amount of change from one drawing to the next is kept small, but gradually increased. This is called ***easing in***. The time spacing between frames is increased so that greater changes in position occur as the object moves faster. When the motion is underway, the changes from frame to frame are held constant. When the motion ends, it is often stopped gradually, by reducing the amount of change from frame to frame of the moving object. This is called ***easing out***. The speed control facility of such non-linear animation sequences can be effectively used while it is required to synchronize animation with audio playback.

Note that spline interpolation doesn't mean that objects should follow Bezier shaped paths, but the rate at which their properties change should be interpolated using a Bezier function say $f(t)$. Intermediate parameter p_t can be calculated as $p_t = (1 - f(t))\, p_1 + f(t)\, p_2$.

Thus, ***spatial interpolation*** defines the ***motion paths*** or change of object position in space. ***Temporal interpolation***, on the other hand, affects the rate of change of objects position with time. Given the vertex positions at the key frames, we can fit the positions with linear or nonlinear motion paths. Interpolation can be applied to other properties of a layer. Its angle can be varied, so that it appears to rotate. **Zoom in** and **zoom out** effect, i.e., the impression of approaching or receding movement can be brought about by scaling or interpolating size.

For instance, (see the image given) the flower, the bee and the comb shown in the image are three objects defined for animation. The bee will be animated to appear to fly from the flower to the comb by following a curved path. The path-curve is then defined as motion path with the initial and final positions of the bee as shown in keyframes 1 and 7. The in-betweens frames (2-6) showing intermediate position of the bee along the motion path are generated by the software (Macromedia Flash) itself.



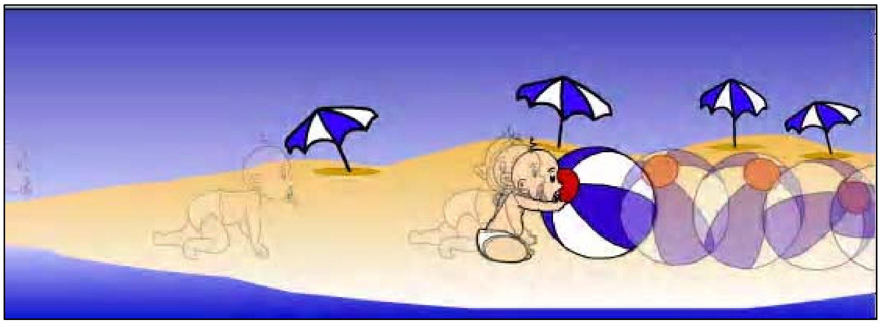


Further to geometrical transformations, parameters for different effects (for example, brightness of *glowing edges*) and filters (for example, radius of ***Gausian blur***) of bitmapped images can be made to vary over time using the standard methods of interpolation between keyframes. Other elements of animation which can be employed for bitmapped images include extensions, tilting, bending, lofting, rendering, fading (in or out) and exploding. These elements change in successive frames as time marches on thereby creating a flowing series of changing imagery what we call as ***motion graphics***. Changes can occur independent of or in concert with other changes. For example, you can make an object rotate and fade in as it moves across.

Besides other special effects the camera panning effect like in the movies can be simulated in animation. What we call ***panning*** is the familiar sweep of the background across the field of view as the movie camera turns. Frequently, a background drawing

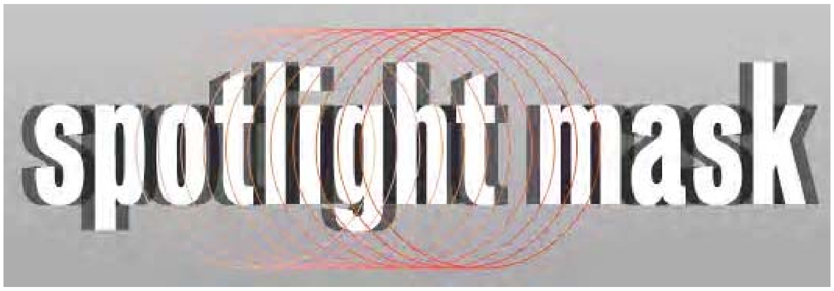
NOTES

larger than the field of the camera, is moved step by step across the animation table as the camera exposes frame after frame of film. Objects in the foreground appear to be moving along relative to the scenery behind them. Thus, we can make a bird fly while the background (sky) rolls in opposite direction, giving the illusion that the bird is covering distance relative to the background.



In the above frame, the baby crawls from left to right and then holds the ball with his hands. The ball spins and rolls from right to middle of the scene





The above frame shows the Spotlight passes over the text from left to right – such animation effect is often used in title casting of motion pictures

3D Animation

A distinct alternative to the 2D animations techniques discussed so far is 3D animation or ***stop-motion*** **animation**.

The classical technique is to create 3D object models out of malleable modeling material, such as plasticine and manipulate those objects in 3D miniature sets between shots to produce natural movement, gesture and otherwise impossible changes. This form of animation is often called ***clay animation***.

NOTES

In the digital realm, 3D wireframe models are created first and then surface and material properties are assigned using photo-realistic rendereing. There are distinct numerical parameters that control objects position (movement) and orientation (rotation) in space, its surface characteristics, its shape, intensity and direction of light sources, and camera position and angle. A 3D animation is achieved by rendereing a scene as the first frame, make some changes to the parameters, render the next frame, and so on. Motion paths in 3D (often 3D Bezier splines) can be used to describe movement.

Realistic shading and rendering based on advanced ray tracing algorithm consumes considerable time to generate a scene. Therefore, high processing power and memory is required to cope up with the required frame rate for smooth animation.

At the very highest level of 3D computer-generated animation software interfaces allows the animator to control different movement parameters to produce smooth movement across the frames. Described below briefly are the different methods of controlling animation.

(i) ***Full Explicit Control*** – It is the simplest type of control where the animator either specifies simple changes like scaling, translation, rotation, or provides keyframe information and interpolation methods interactively.

(ii) ***Procedural Control*** – It is based on certain kinds of behavior that can be applied to objects and the way they interact. In a ***physically based system*** the position of one object may influence the motion of another object (for example, spotlight follows a dancer, a sunflower follows the sun). In such systems objects are modeled with physical attributes such as mass, moment of inertia, elasticity, velocity etc. and object behavior as emulated in animation are based on laws of Newtonian Physics against applied external force. Thus moving objects can be made to collide realistically, or bounce of solid surface.

Procedural control: It is based on certain kinds of behavior that can be applied to objects and the way they interact

(iii) ***Kinematics*** – It is the study of motion of bodies without reference to mass or force. That is it is only concerned with how things can move, rather than what makes them do so. Animations of linked objects or jointed structures (for example, limbs of human or animal figures) are controlled by imposing *kinematic constraints* obeyed by real objects or structures. For example, a 3D model of a door must have the same ***degree of freedom*** to move/rotate as a real door has with the movement constraints produced by the hinges.

Kinematics being a general term, ***forward kinematics*** and ***inverse kinematics*** – both are used in controlling animation. While the former deals with linked motions from cause to the effect, the inverse kinematics works backward from effect to cause. For example, it is the motion of the upper arm that propels the rest of the arm and hand. Modeling the hands position from movement and position of the upper arm requires forward kinematics. Whereas, first fixing the position of the hand and then backtracking to find the relevant motion of upper arm is what inverse kinematics is and sometimes it is more useful to the animator.

(iv) ***Tracking Live Action*** – This technique produces exceptionally realistic motion. Trajectories of objects to be animated can be generated by tracking

of live action. One such method is ***rotoscoping***. A film is made in which people or animals act out the parts of the characters in animation. Then the animator draws over the film, changing the background and replacing the human or animal actors with their animation equivalents. In an alternative method some sort of indicators or motion sensors are attached to key points on an actor's body or body suit. By tracking the position of the indicators or sensors, the animator can get locations for corresponding key points in an animated model.

NOTES

2.4 SPEECH RECOGNITION

Speech generation and recognition system enables human beings to communicate with the system verbally. Commands that human beings earlier used by inputs through keyboard or mouse using hands (fingers) and eyes can now be issued by using speech that the computer recognizes. But to accomplish this, computers must have the ability to recognize the voice or speech to execute the command. To communicate efficiently, computers must have speech recognition software that can interpret human speech.

Such capability enables a person to keep his or her hands and eyes free. A person may dictate the computer to write for him and act like a steno-typist. This can be done in real mode in which a person is dictating by using a microphone and the computer is executing the command directly and live. Such dictations can be stored on the computer in the form of some audio files and the same thing can also be done in batch mode later, as per convenience. This is also a multimedia application. By using this system, matter preparation becomes easy. Users of these systems can do other things that require the use of hands and eyes while his mouth and ears instruct computers to do things for him. Incorporating this system in the computer will enable human beings the to do things in a more convenient way.

Cell phones that recognize human voice to dial numbers are already in the market. Text readers have also come to the market. Software performing the task of a dictaphone that can instruct computer to type matters as spoken has also appeared. Optical recognition software packages are already popular. So, now, it is time for voice recognition software to be popular. Speech recognition and generation can be automated by creating software that works as an audio interface.

Check Your Progress

1. What is a font?
2. Define the term kerning.
3. What is the use of data compression?
4. What do you mean by formatted text document?
5. Define the term digital image.
6. List the commonly used vector file formats commonly used in IT industries.
7. What is visible light?
8. How is image acquisition done?
9. List the common compressed file formats.
10. What is a period?
11. Define the term wavelength.
12. What is the range of human hearing?
13. What is pitch?
14. What are the elements of an audio system?
15. What is a microphone?

Speech Recognition

Generation of speech is easier in comparison to the recognition of speech. It is much more difficult to recognize than to generate speech. This function of speech recognition is the task that is performed well by the human brain. Computer on other hand, performs this task poorly. This lack of performance is due to the fact that till date such a remarkable software has not been developed. But it is possible to design such as software that performs many functions that human brain performs so easily. Comprehending some idea by the computer requires different types data of vast amount. This is possible since digital computers are capable of storing and retrieving or recalling huge data volume at extremely high speeds. Further, they can perform high-speed mathematical calculations and repetitive tasks without getting bored or fatigued.

Poor performance of computer is found when raw sensory data is present to it. One can program a computer to send a telephone bill but to program it, to teach it to understand a human voice is a major issue related to voice recognition. Voice signals after being digitized, are processed.

Digital signal processing recognizes voice in two stages:

1. Extraction and matching of feature. Every word of audio incoming signal is analyzed in isolation for identifying resonate frequencies and excitation type.
2. Such parameters go under comparison subsequently with earlier spoken words as examples for identifying the closest match.

This system has limitations of words to few hundred words only and unable to understand if spoken in quick succession. For this system to accept speech, it should be spoken with distinct pause. Further, the computer has to be retrained for every individual speaker.

Speech recognition system basically contains the following essential parts:

(i) Voice input device.

(ii) First storage device containing first recognition word indicating the pronunciation of a word that would undergo speech recognition.

(iii) Generating a device to judge whether a given word contains specific words as a part of the given word.

Speech Generation

To generate speech by computer two approaches are used:

1. Digitization and recording.
2. Simulation of vocal tract.

The first approach digitizes the voice of a human speaker and then stores it in a compressed form. This digitized and stored voice is played back after uncompressing the stored data and converting it back to an analog signal that is accepted by the speaker.

Simulation of vocal tract is quite complicated. It mimics the real physical mechanisms of speech creation in human beings. Vocal tract can be viewed as an acoustic cavity for human beings that resonate at some frequencies which depends on the constructional features of chamber such as shape and size.

2.5 EXTENDED IMAGES

Extended images, such as route panoramas, scene tunnels, panoramic views and spherical views are acquired in an urban area and associated with geospatial locations. To generate a scanning plan, based on visibility, image properties and importance of scenes, a 3D LIDAR elevation map is used. Scanning scenes along streets and at spots of interest allows the compact and complete visual data collection.

NOTES

Check Your Progress

16. Fill in the blanks with appropriate words.
    a. Images can be generated and stored in the PC in typically two different ways. One is called vector art and the other is referred to as _________.
    b. The _________ protocol defines standard multibyte messages or instructions to control some aspect of the performance of an instrument.
    c. _________ are image frames that depict the key-positions of the objects being animated and marks significant changes in the animation sequence.
    d. _________ is a document-layout and hyperlink-specification language that is used to create hypertext documents and web pages.
17. State whether the following statements are true or false.
    a. In Huffman coding the frequency of each character in the text file is analysed for the encoding operation.
    b. An image is not a very important component of digital multimedia. It is the representation of an object or a two- or three-dimensional scene on a planer region (spatial representation).
    c. In digital system, colour depth or pixel depth refers to the number of bits associated with each pixel in a bitmap.
    d. Most high-end professional sound cards do not have the built-in amplifiers, because they cater to amplified speakers.

To present spaces, maps/satellite images which are mostly used and from which snapshots and spherical images have been indexed to wide and open locations, the extended images and used. The goals of extended work is to acquire real scenes in areas using extended image media, such as route panoramas, scene tunnels, spherical views and digital images and realize Web based geospatial information visualization.

Though there is slight 2D distortion in the extended images for their specific projections, their features are:

- **Compact:** It includes less redundant scenes than video.
- **Complete:** It covers every street and many other locations on the map.
- **Continuous:** In it, the long route scenes are suitable for navigation.
- **Comprehensive:** Indexes from (or to) maps and spaces provide flexible space transition and traversing in the media.

2.6 DIGITAL INK

In multimedia note-taking systems, interacting with data, such as digital ink and audio is challenging. Dynamically grouping digital ink and audio is required to support user interaction in freeform note-taking systems. For audio, groups might be a complete spoken phrase or a speaker turn in a conversation. Digital ink and audio grouping is important for editing operations, such as deleting or moving chunks of ink and audio notes.

Groups of ink or audio in range of sizes, depending on the level in the hierarchy is yielded by clustering algorithm. It thus provides structure for simple interactive selection and rapid non-linear expansion of a selection. Examples of such ink and audio note-taking systems are Filochat, Dynomite etc. Whenever the user wishes to edit or browse the ink or audio data there is a lack of structure which becomes limiting feature for the modification and thus bounds the user from doing any kind of changes in the system. The challenge is to provide computational support to dynamically group the ink and audio data into units that are useful for user interaction.

For digital ink and audio, these computations are far more complicated than the ones for online text. In online text, selection of a particular word or line is easy. On the other hand, selection of handwritten ink corresponding to a word or line is difficult. This is because digital ink is composed of strokes which are graphical objects in a freeform space and not automatically mapped into meaningful units.

Even when digital ink is mapped onto a timeline, it is difficult to select a segment of audio corresponding to a phrase or a turn in a conversation. These characteristics of ink and audio make it tedious for the user to manipulate multimedia data.

2.7 SUMMARY

In this unit, you have learnt that:

- Text is a fundamental building block in a multimedia system. It is one of the most widely used and flexible means of presenting information and conveying ideas in a multimedia environment.

NOTES

- An alphabet is a complete standardized set of letters—the basic written symbols to communicate in a particular language.
- The plain or unformatted text is the most elementary fixed size character sets. The .txt file created using notepad is an example of plain text.
- In order to represent text in a digital form, each character of a particular language has to be related to a specific bit pattern.
- In a formatted text, control characters manage the appearance of the text. As a result, you can make a string of text appear in any combination of bold, underlined, italic, paragraphed and tabulated style.
- The control characters used in the application software may vary. So the appearance of a document created using MS Word may look different in an HTML document and vice versa.
- Hypertext is a special type of formatted text. In the context of text being used as the fundamental building block of multimedia applications, the powerful processing capabilities of a computer can be applied to make the text more interactive and organize the content in non-sequential way.
- By positioning the mouse pointer on a portion of a text (a word or even a paragraph on the screen called anchor) and then clicking, you may jump to the linked destination and display multimedia information (text, image, video, etc.) in the same screen or on another screen.
- Hypertext is a special text format that is used to link multimedia information in a non-sequential way.
- The meaning of the word 'hyper' is something close to 'extra' or 'beyond'.
- HyperText Markup Language (HTML) is a document layout and hyperlink specification language that is used to create hypertext documents and web pages.
- Basically, HTML files are just plain ASCII text files that can be created in any standard word processors even in Windows Notepad. Such files contain two things—the normal textual content and the markup 'tags'.
- These tags are HTML instructions written within '<', '>' symbols specifying the presentation format (such as size, font, colour, location, etc.) of the textual content. The markup tags are usually paired with an ending tag starting with a slash (`</[tag]>`).
- Tags can be used to establish hyperlinks to documents, image files, music files, Java applets, etc., from within the document.
- If the HTML file contains a `<a href>` tag, the browser knows that what follows describes a hyperlink to another document.
- HTML presumes a Document Type Definition (DTD), which specifies valid tag names, attributes and their syntax.
- A font is a collection of characters of a specific style and size of a particular typeface. For example, Times New Roman is a typeface, and you may choose different fonts (having specific style and size) from within it.

NOTES

- Typefaces are the shapes or graphic representations of the characters, numbers or special characters that are stored internally in the computer as bits.
- There are basically two types of typefaces—and thus two types of fonts—Serif and Sans serif.
- The vector fonts draw the characters by using vector drawing primitives using mathematical functions, thus requiring considerable smaller size than the bitmap fonts.
- Compressed data occupies less space for storage and also takes less time for communication. The data may be text, image, audio, video or animation objects.
- There are fundamentally two types of data compression—lossy and lossless.
- Huffman coding functions by analysing the relative frequency of occurrence of different characters in a text file.
- The characters in the text file that have the highest frequency of occurrence are assigned the shortest encoding with the fewest bits. Characters with lower frequencies get assigned longer encoding with more bits. Thus, compression is achieved by overall saving in the total number of bits.
- In Huffman coding the frequency of each character in the text file is analysed for the encoding operation.
- An image is a very important component of digital multimedia. It is the representation of an object or a two- or three-dimensional scene on a planar region (spatial representation).
- Images can be generated and stored in a personal computer in two typically different ways. One is called vector graphics and the other is referred to as bitmapped.
- A piece of vector art is a file that contains descriptions of how to generate the image but not the actual image itself.
- PostScript files, developed by Adobe, are generated by DTP packages and authoring systems while WMF was developed by Microsoft and it is an excellent format for image interchange between Windows applications.
- HPGL is an interpreted vector description language meant for plotters, and DXF is the most widely accepted format for interchange of engineering graphics data between different CAD packages, such as AutoCAD, etc.
- A bitmapped image, in contrast, has in the file the actual pixel image data. That is, it simply holds the color number for each dot or pixel in an image.
- JPEG, on the other hand, is an example of a lossy compression—data from the original image, which is deemed to be redundant, is thrown away in the compression process.
- Light is an electromagnetic wave. The human eye is able to 'see' only a very small part of the total range of electromagnetic radiation called the visible light.

NOTES

- Light waves having a single wavelength are called monochromatic light, and the colors produced by a visible light of a single wavelength are pure spectral colors. The light produced by a laser torch is monochromatic, having a single wavelength.
- One approach to create a broad range of colors is by suitable combinations of three primary colours.
- The CMY (Cyan, Magenta and Yellow) model is widely used in professional four color printing processes. The colour printer is programmed to combine different amounts of cyan, magenta and yellow inks to create a color.
- Digital image processing refers to processing of digital images by means of a digital computer where the input is an image and the output is another image or a set of characteristics or parameters acquired from the image.
- TWAIN is an image capture API, developed by consortium of Hewlett-Packard, Kodak, Aldus, Logitech and Caere for Microsoft Windows and Apple Macintosh operating systems.
- The API uses a four-layer protocol (device layer, acquisition layer, protocol layer and application layer) for connecting TWAIN compliant devices with TWAIN compliant applications, mostly through USB interface.
- The Image and Scanner Interface Specification (ISIS) has more functions than TWAIN and is mainly used with the SCSI-2 interface.
- Bitmaps are generally created by scanners, digital cameras and paint programs, such as Corel Paint Shop Pro®, (image processing programs) Adobe Photoshop®, etc.
- Literally speaking, to animate is to bring to life, i.e., to put something into action. Animation makes graphics more realistic by imparting motion and dimension to an inanimate object. A computer animation sequence can be set up by specifying the storyboard, the object definitions, and the image frames. The storyboard is an outline of action. It could consist of rough sketches of motion sequence or it could be a list of basic events that are to take place.
- Object definitions are given for each participating object in terms of their shape and movement.
- Classically, picture frames depicting animated sequence were drawn manually. In doing so the onionskin technique which is popularly known as cel animation technique is mostly adopted.
- Keyframes are image frames that depict the key-positions of the objects being animated and marks significant changes in the animation sequence.
- Vector graphic images, on the other hand, are created using graphics primitives, such as lines, arcs, etc., governed by mathematical equations describing the shapes and colors are applied to those primitive shapes.
- Image size is the physical dimensions of an image when it is printed out or displayed on a computer screen. It is normally expressed in inches or centimetres.

NOTES

- In digital system, color depth or pixel depth refers to the number of bits associated with each pixel in a bitmap.
- When sound travels through air, alternating high and low pressure are created along the wave path.
- The pressure amplitude of a sound wave measures the change in sound pressure inside the wave.
- The distance between two successive crests is known as the wavelength. It is the distance that a wave travels in the time of one oscillatory cycle.
- Pitch is the characteristic of a musical sound by which human ears differentiate between a shrill sound and a dull sound.
- In other words, pitch is a measure of the frequency of a sound wave. A shrill or high-pitched sound has higher frequency than a flat or dull sound.
- In music, musical notes are ordered from low pitch to high pitch forming a scale that provides the basis for a musical composition.
- Timbre is the distinctive quality or tone of a sound that may come from a singing voice or a musical instrument.
- The perception of loudness of a sound to the human ear depends mainly upon the pressure amplitude, but it also depends upon duration, frequency, presence or absence of background noises, as well as the sensitivity of the ear.
- Masking occurs when one sound prevents us from hearing a second sound. When you hear a loud sound and a soft sound simultaneously, your ears receive both the sound signals but our brain ignores the soft one and as a result you cannot 'hear' the soft sound. This phenomenon is known as masking.
- The phenomenon of masking is one of the most important one in psychoacoustics study, since it gives us a basis for eliminating sound information that is not perceived anyway. It helps a great deal in compressing digital audio files.
- The microphone is a device used to capture audio waves and convert to electrical signal.
- Dynamic or moving coil microphones work by vibrating a thin metallic diaphragm and an attached coil of wire in a magnetic field. Generally, the thin diaphragm is attached to a coil of wire that surrounds or is surrounded by a high-powered magnet.
- Analog-to-Digital Converter (ADC) translates the analog sound waves into digital data that the computer can recognize.
- The DSP (Digital Signal Processing) determines how many Musical Instrument Digital Interface (MIDI) channels, sound streams or voices the sound card can support.
- Most high-end professional sound cards do not have the built-in amplifiers, because they cater to amplified speakers.

- MIDI is a standardized protocol or procedure set. Manufacturers of musical instruments, computers and computer software now routinely adopt MIDI protocol.
- This drawback of MIDI has been corrected to some extent by including additional information that can be included in the MIDI event, such as how a note should be modulated or bent.
- The term 'computer graphics' was first used by William Fetter in 1960. Fetter was a graphic designer for Boeing Aircraft Co.
- A pixel (also called picture element) can be defined as the smallest piece of information in an image.
- Spatial interpolation defines the motion paths or change of object position in space. Temporal interpolation, on the other hand, affects the rate of change of objects position with time.
- In the digital realm, 3D wireframe models are created first and then surface and material properties are assigned using photo-realistic rendering.
- Speech generation and recognition system enables human beings to communicate with the system verbally. Commands that human beings earlier used by inputs through keyboard or mouse using hands (fingers) and eyes can now be issued by using speech that the computer recognizes.
- Generation of speech is easier in comparison to the recognition of speech. It is much more difficult to recognize than to generate speech. This function of speech recognition is the task that is performed well by the human brain.
- Extended images, such as route panoramas, scene tunnels, panoramic views and spherical views are acquired in an urban area and associated with geospatial locations.
- In multimedia note-taking systems, interacting with data such as digital ink and audio is challenging.
- Dynamically grouping digital ink and audio is required to support user interaction in freeform note-taking systems.
- Digital ink and audio grouping is important for editing operations, such as deleting or moving chunks of ink and audio notes.

2.8 ANSWERS TO 'CHECK YOUR PROGRESS'

1. A font is a collection of characters of a specific style and size of a particular typeface. For example, Times New Roman is a typeface and you may choose different fonts (having specific style and size) from within it, as follows:
    - Times New Roman 14 point Italic – one font
    - Times New Roman 14 point Bold – another font
    - Times New Roman 12 point Bold and Italic – yet another font.

NOTES

2. Kerning is the horizontal spacing between characters. In many word processing software the kerning can be increased or decreased to make the letters in the words more spread out or compact.
3. Data compression brings down the size of a digital data file and compressed data occupies less space for storage and also takes less time for data communication. The data can be text, image, audio, video or animation objects.
4. The formatted text document created using Microsoft Word or Wordpad packages have by default the .doc extension. This format is very common and has a rich set of formatting features. It also supports images and graphics. Most open source word processing software these days supports this format.
5. A digital image can be considered as a set of picture elements (pixels). These pixels are like the tiny dots or pigments on a photograph printout arranged in rows and columns that makes up the image. Each pixel corresponds to a color value at a particular portion of the image.
6. Some of the vector file formats commonly used in the IT industries are:
    - PostScript file.
    - Computer Graphics Metafile (*.CGM).
    - Windows Metafile (*.WMF).
    - Hewlett Packard Graphics Language or HPGL (*.PLO).
    - Data Exchange Format (*.DXF).
7. Light is an electromagnetic wave. The human eye is able to ‘see’ only a very small part of the total range of electromagnetic radiation called visible light.
8. In image acquisition, image may be acquired from analog source, such as a conventional photograph or a frame of a video clipping, or it may be available in digital form like a photograph taken with a digital camera. Generally, digital cameras and scanners are the input devices for digital input.
9. The most common compressed file formats are:
    - Graphics Image Format (*.GIF).
    - Tagged Information File Format or TIFF (*.TIF).
    - Joint Photographic Experts Group or JPEG (*.JPG).
    - Windows Bitmap (*.BMP) and Windows Device Independent Bitmap (*.DIB).
10. The amount of time taken by a wave to complete one cycle is the period of the wave. Period and frequency are reciprocals of each other.
11. Wavelength is the distance between two successive crests and is the distance that a wave travels in the time of one oscillatory cycle.
12. The range of hearing for a healthy young person is 20 to 20,000 Hertz, i.e., 20 KHz. This range of frequency is called the audible range. This audible range is kept in mind when creating multimedia audio.

13. Pitch is the characteristic of a musical sound by which human ears differentiate between a shrill sound and a dull sound.
14. The minimal hardware set-up for digital audio recording consists of the following: Microphone à Amplifier → A/D Converter → Storage Device (HDD or DAT) → D/A Converter → Speaker.
15. Microphone is a device used to capture audio waves and convert to electrical signal. It is the first component in the chain of components used in a digital audio studio, and in many ways the most important. The microphone is a transducer or a device that converts mechanical energy (travelling waves of compression and rarefaction in a medium like air) into patterns of electrical current.
16. (a) Bitmapped; (b) MIDI communication; (c) Keyframes; (d) HyperText Markup Language (HTML)
17. (a) True; (b) False; (c) True; (d) True.

NOTES

2.9 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is a hypertext?
2. What is text compression?
3. What are the formats in which files can be formatted?
4. How does the CMY color model work?
5. What is TWAIN?
6. What do you understand by color depth?
7. What are the files formats in images?
8. What is meant by audio?
9. Write short notes on pitch, timbre and loudness.
10. What is MIDI?
11. How does the analog video camera work?
12. How does the transmission of video signal happen?
13. What is animation?
14. What does speech recognition system basically contain?

Long-Answer Questions

1. Explain the various types of images.
2. Explain the working of the RGB color model.
3. Discuss the steps involved in image processing.

NOTES

4. Explain the significane of resolution.
5. What do you understand by color management policy?
6. Distinguish between pure tones and note.
7. What is the dynamic range of human hearing? Discuss.
8. Explain how masking can be done.
9. What does speech recognition system basically contain? Describe with the help of example.
10. Explain the significance of speech recognition.

UNIT 3 DIGITAL VIDEO AND IMAGE COMPRESSION

NOTES

Structure

3.0 INTRODUCTION

In this unit, you will learn about the significance of digital video and image compression. Compression is a reversible process of conversion of data to a format, which requires fewer bits, so that the data can be stored or transmitted more efficiently. Compression algorithms are divided into two fundamental types (i) Lossless compression and (ii) Lossy compression. In a lossless compression, no data or information is lost at the time of compression and decompression process. While compression condenses the size of the file, the decompression process restores the data back in its original value and size. Lossy compression, usually applied to image data, does not allow reproduction of an exact replica of the original image. Thus, lossy compression allows only an approximation of the original to be generated.

Generally some elements within the data are more common than others and most compression algorithms exploit this property, known as redundancy. The greater the redundancy within the data, the more successful the compression of the data is likely to be. Fortunately, digital video contains a great deal of redundancy and thus is very suitable for compression. You will also learn about the significance of redundancy and visibility. Finally, you will learn about the significance and the various standards of image and video compression. Various algorithms for compression is also being discussed in this unit.

NOTES

3.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Learn about the basics of compression and compression system
- Evaluate a compression system
- Discuss the significance of redundancy and visibility
- Explain various video compression techniques
- Describe the various video compression techniques
- Describe image compression standards

3.2 INTRODUCTION TO COMPRESSION

You have studied that for good reproduction of the original analog signal, it is necessary that the text, image, sound or animation sequence be digitized at appropriate resolution and quantization levels. This process of sampling and quantization generates millions of samples, each sample ca have several bytes in length (depending on whether the sampled signal is text, image or audio). So, digital multimedia object files are normally very large. For the purpose of storage and transmission, they are required to be compressed. Otherwise, even with fast Internet access and availability of large and affordable storage devices, the handling of such big digital media files would have been prohibitive. On the other hand, it is not desirable to forgo the quality of the digital images, audio or video in the process of compression. Compression algorithms are divided into two fundamental types (i) Lossless compression and (ii) Lossy compression. In a lossless compression, no data or information is lost at the time of compression and decompression process. While compression condenses the size of the file, the decompression process restores the data back in its original value and size. On the other hand, lossy compression sacrifices some information. However, the information that is sacrificed utilizing the limitations of human vision or hearing and the loss of fidelity is not perceptible to a human being. For example, for a sound file, it may be the frequencies that are inaudible to the human ear; or a low amplitude sound immediately after a high amplitude sound that is imperceptible. In case of an image file, it may be slight variations in color that the eye cannot detect.

Apart from the broad categories of lossless and lossy compression, the compression algorithms may be tagged based on the techniques adopted, such as, entropy, dictionary-based, arithmetic, adaptive, perceptual and differential compression methods.

The *compression rate* of a specific compression algorithm is the ratio of the original file size to the size of the compressed file. Often it is expressed as a percentage. For example, if an image or audio file is reduced by compression to 25 per cent of its original size, it can be said that 75 per cent compression is achieved. Alternatively, it can also be said that the compression rate is 4:1.

The audio and image compression methods will briefly be reviewed.

3.3 EVALUATING A COMPRESSION SYSTEM

NOTES

Multimedia data compression technique is used to reduce the redundancies in data representation with reference to decrease in the data storage requirements and hence communication overloads when transmitted through a communication network. If the compressed data are properly indexed then it improves the performance of mining data in the compressed large database as well. This is particularly useful when interactivity is involved with a data compressing system. For example, image compression for Joint Photographic Experts Group (JPEG) JPEG2000 files and video compression is used for Moving Picture Experts Group (MPEG), MPEG-4/7 files and image mining is used for Content Based Image Retrieval (CBIR), video event detection, etc. JPEG 2000 is an image compression standard and coding system. The essential feature of data compression is known as the 'compression ratio'. This ratio refers to the size of a compressed file to the original uncompressed file. For example, suppose a data file takes up 100 kilobytes (KB). Using data compression software, that file could be reduced in size for example 50 KB which makes easy to store on disk and faster to transmit over an Internet connection. In this specific case, the data compression software reduces the size of the data file by a factor of two or results in a compression ratio of 2:1. There are 'lossless' and 'lossy' forms of data compression work in the field of preparing multimedia applications and Web designing. Lossless data compression is used when the data has to be uncompressed exactly as it was before compression. Text files are stored using lossless techniques, since losing a single character in the worst case can make the text dangerously misleading. Archival storage of master sources for images, video data and audio data generally needs to be lossless state. However, there are strict limitations to the amount of compression that can be obtained with lossless compression. Lossless compression ratios are generally in the range of 2:1 to 8:1. Lossy compression, in contrast, works on the assumption that the data does not have to be stored perfectly. Much information can be simply thrown away from images, video data and audio data and when uncompressed such data will still be of acceptable quality. Compression ratios can be an order of magnitude greater than those available from lossless methods. Each has its own uses, with lossless techniques better in some cases and lossy techniques better in others. In fact, as this document will show, lossless and lossy techniques are often used together to obtain the highest compression ratios. Even a specific type of file, the contents of the file, the orderliness and redundancy of the data can strongly influence the compression ratio. In some cases, using a particular data compression technique on a data file where there is not a good match between the files can actually result in a bigger file. The two prime types of data compression done with reference to Web designs and development phase are known as audio compression and video compression. Compressing an image is significantly different than compressing raw binary data. The general purpose compression programs can be used to compress images but the result is less than optimal. This is because images have certain statistical properties which can be exploited by encoders specifically designed for them. Some of the finer details in the image can be reduced for saving the file which exploits bandwidth or storage space. This also means that lossy compression techniques can be used in this area. Lossless compression involves with compressing data which when decompressed will be an exact replica of the original data. This is the case when binary data, such as executables, documents,

NOTES

etc., are compressed. They need to be exactly reproduced when decompressed. On the other hand, images and music too need not be reproduced exactly. An approximation of the original image is enough for most purposes, as long as the error between the original and the compressed image is presentable. Figure 3.1 demonstrates JPEG/JPEG2000 compression of gray scaled image.

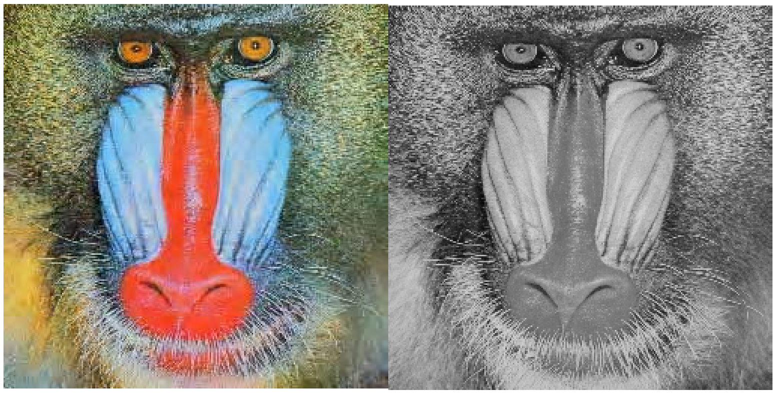


Fig. 3.1 *Original and Compressed Gray Scaled Image*

Error metrics is used as prime tool to compress image and data. The two types of the error metrics used to compare the various image compression techniques are the Mean Square Error (MSE) and the Peak Signal to Noise Ratio (PSNR). The MSE is the cumulative squared error between the compressed and the original image, whereas PSNR is a measure of the peak error. The mathematical formulae for calculating MSE and PSNR are as follows:

$$\text{MSE} = \frac{1}{\text{MN}} \sum_{y=1}^{1} \sum_{x=1}^{N} [\text{I}(x, y) - \text{I}'(x, y)]^2$$

$$\text{PSNR} = 20 \times \log10\,(255 / \text{sqrt(MSE)})$$

Where, I(x,y) is the original image, I′ (x,y) is the approximated version which is actually the decompressed image and M, N are the dimensions of the images. A lower value for MSE means lesser error and as seen from the inverse relation between the MSE and PSNR, this translates to a high value of PSNR. Logically, a higher value of PSNR is good because it means that the ratio of Signal to Noise is higher. Signal is the original image and noise is the error in reconstruction. So, if you find a compression scheme having a lower MSE and a high PSNR you can recognize that it is a better one. Compressing gray scale images refer to the algorithms explained can be easily extended to color images either by processing each of the color planes separately or by transforming the image from RGB representation to other convenient representations like YUV in which the processing is much easier. YUV is a color space typically used as part of a color image pipeline. The YUV color space (color model) differs from RGB which is what the camera captures and what humans view. Following functions are performed in compressing the images and data:

- Specifying the Rate (bits available) and Distortion (tolerable error) parameters for the target image.
- Dividing the image data into various classes based on their importance.

- Dividing the available bit budget among these classes such that the distortion is a minimum.
- Quantize each class separately using the bit allocation information derived in previous step.
- Encode each class separately using an entropy coder and write to the file.

In the fractal image compression technique, where possible self similarity within the image is identified and used to reduce the amount of data required to reproduce the image.

3.4 REDUNDANCY AND VISIBILITY

Redundancy in a digital video image occurs when the same information is transmitted more than once. For example, in any area of the picture where the same color spans more than one pixel location, there will be redundancy between pixels, since adjacent pixels will have the same value. This will apply to horizontal and vertical representation.

Other example could be when the scene or part of the scene contains predominantly vertical oriented objects then there is a possibility that two adjacent lines will be partially or completely redundant at the same. These two types of redundancy (pixel and line) exist in any image and are called spatial redundancy.

Compression techniques take advantage of redundancy or coherence in images, here are two types of redundancy:

- **Spatial:** Its key points are:
  - o Pixels next to each other on the scanline are the same or close in value
  - o Adjacent scanlines are completely or partially the same, with small differences
- **Temporal:** Its key points are:
  - o Adjacent frames may be the same, or related by a simple transformation
  - o Moving objects yield small changes frame-to-frame

3.5 VIDEO COMPRESSION TECHNIQUES

Out of different multimedia elements the need for compression is greatest for video as the data volume for Full Screen Full Motion (FSFM) video is very high. Frame size for NTSC video is 640 pixels by 480 pixels and if we use 24 bits color depth then each frame occupies 640 × 480 × 3 bytes, i.e., 900 KB. So each second of NTSC video comprising 30 frames occupies 900 × 30 KB which is around 26 MB and each minute occupies 26 × 60, i.e., 1.9 GB. Thus a 600 MB CD would contain maximum 22 seconds of FSFM video. Now imagine the storage space required for a 2-hour movie. So the only way to achieve digital motion video on PC is to reduce or compress the redundant data in video files.

Redundancy in digital video occurs when the same information is transmitted more than once. Primarily in any area of an image frame where same color or intensity spans more than one pixel location, there is ***spatial redundancy***.

NOTES

Redundancy: In a digital video image, it occurs when the same information is transmitted more than once

Check Your Progress

1. What are two types of compression algorithm?
2. What do you mean by multimedia data compression technique?
3. A when does redundancy in multimedia occurs.
4. What is temporal redundancy?
5. What is lossless compression?
6. Define the term run-length encoding.
7. How transform coding works?
8. What does MPEG stands for?
9. How DVI-A differs from DVI-I?
10. Give any one use of DVI technology.

NOTES

Secondly when a scene is stationary or only slightly moving, there is redundancy between frames of motion sequence – the contents of consecutive frames in time are similar, or they may be related by a simple translation function. This kind of redundancy is called ***temporal redundancy***.

Spatial redundancy is removed by compressing each individual image frame in isolation and the techniques used are generally called ***spatial compression*** or ***intra–frame compression***. Temporal redundancy is removed by storing only the differences of subsequent frames instead of compressing each frame independently and the technique is known as ***temporal compression*** or ***inter-frame compression***.

Spatial compression applies different lossless and lossy method same as those applied for still images. Some of these methods are:

(*i*) Truncation of least significant image data.

(*ii*) Run Length Encoding or RLE.

(*iii*) Interpolative techniques.

(*iv*) Predictive technique–**DPCM** (Differential Pulse Code Modulation), **ADPCM** (Adaptive DPCM).

(*v*) Transform Coding Techniques–**DCT** (Discrete Cosine Transform).

(*vi*) Statistical or **Entropy Coding**–**Huffman Coding**, LZW coding, Arithmetic coding.

The most simplistic approach for temporal compression is to perform a pixel-by-pixel comparison (subtraction) between two consecutive frames. The compare should produce zero for pixels, which have not changed, and non-zero for pixels, which are somehow involved in motion. Then only the pixels with non-zero differences can be coded and stored, thus reducing the burden of storing all the pixel value of a frame. But there are certain problems with this approach. Firstly even if there is no object motion in a frame, slightest movement of camera would produce non-zero difference of all or most pixels. Secondly quantization noise would yield non-zero difference of stationary pixels.

In an alternative approach the motion generators camera and/or object can be 'compensated' by detecting the displacements (***motion vectors***) of corresponding pixel blocks or regions in the frames and measuring the differences of their content (***prediction error***). Such approach of temporal compression is said to be based on ***motion compensation***. For efficiency each image is divided into ***macroblocks*** of size N × N. The current image frame is referred to as *target frame*. Each macroblock of the target frame is examined with reference to most similar macroblocks in previous and/or next frame called ***reference frame***. This examination is known as ***forward prediction*** or ***backward prediction*** depending on whether the reference frame is a previous frame or next frame. If the target macroblock is found to contain no motion, a code is sent to the decompressor to leave the block the way it was in the reference frame. If the block does have motion the motion vector and difference block need to be coded so that the decompressor can reproduce the target block from the code.

3.5.1 Compression/Decompression (CODEC)

So far you have learned about the various types of compression techniques and the motives behind them. Actually, compression algorithms are executed in a vast range of

methods that are fine tuned and optimized for effective performance in real life. Many of these algorithms are standardized and improved upon by official committees to maintain portability and uniformity across different hardware and software platforms. Also, quite often, instead of using a single algorithm, different compression methods are applied in combination with each other. For example, the MPEG and JPEG compression combine the run length encoding, DCT, as well as the Huffman encoding for compression of photographic images. Often such standardized algorithms are patented and available commercially. Sometimes, they are available free of cost as a shareware. An example of such standardized compression algorithms is the family of MPEG algorithms. Similarly, arithmetic encoding is an image compression algorithm that is covered by patents.

Digital audio/video compression is achieved through precise implementations of compression algorithms called CODECs. *CODECs* are compression/decompression algorithms (CODEC → short for **Co**mpression/**Dec**ompression) that take the digital audio/video files as input and produce compressed versions of the audio/video files at the output. These compressed digital data files can be stored for later decompression and playback. During playback, the compressed files are again converted back to digital audio/video signals (decompression) by the same CODEC software. So the computer where you want to playback the compressed audio/video file must have the same CODEC that was used to compress the file.

A particular CODEC uses a particular audio/video compression algorithms and the compressed file is given a unique extension to specify which CODEC has been used.

Lossless and Lossy CODECs

CODECs are of two types (i) Lossless CODECs and (ii) Lossy CODECs.

When a digital data is compressed using a CODEC, often decompression of the compressed data does not accurately reproduce the same audio/video waveform or in other words, the same sound/video cannot be played back exactly. Thus, the original audio or image signal is changed forever. The CODECs that permanently change the digital signal during compression are called Lossy CODECs.

Lossy CODECs

These eliminate redundant or unnecessary information from the digital signal during compression. As a result, the file size is greatly reduced. Now what is redundant information? In case of an audio file, redundant information is the sound whose frequency lies beyond the two extremities (20Hz and 20 Khz). The frequency that people cannot hear, or the audio signals masked by a high volume sound (remember humans cannot hear a low volume sound just after a high volume sound of similar frequency). The lossy audio CODEC keeps track of such audio signals and chops off the redundant information to reduce the file size. Naturally, once the lossy compression transforms the source file, it cannot be brought back to its original form. Lossy compression has become very popular because of its small file size and also very good audio quality. MP3 and Real Audio files are examples of lossy compression. It is easy to transmit MP3 files as streaming audio on the Internet and to store a large number of files in CDs and iPods.

Lossy CODECs: These eliminate redundant or unnecessary information from the digital signal during compression

NOTES

In lossless compression, on the other hand, the original digital data is not changed permanently in any way during the process of compression. So after decompression you get back the audio data in its original form. The CODECs that compresses the audio/video signal *without* any change in the signal after decompression are called lossless CODECs, since no portion of the original data is changed and the quality is maintained exactly at its original level.

Example of lossless compression is removal of silence in a speech file by first noting the period of silence in a speech and then removing 1.41 MBits of data for each second (recall the simple arithmetic: 44100 samples/sec x 16 bits/sample x 2 channels = 1.41 Mbits/sec.) of the silent portion in the speech. During playback, the CODEC will reintroduce silence in the audio and reproduce the exact sound.

Another example of lossless CODECs is the Differential Pulse Code Modulation (DPCM). It is based on storing the difference between consecutive samples instead of storing each of the sampled digital values separately. DPCM will be effective only if the difference between consecutive signals can be stored using lesser number of bits than the signal itself. In this case also, the original values of the sample can be reproduced exactly, the compression type is lossless.

Various algorithms exist for compression of digital audio data, including Huffman algorithm, ADPCM, µ-law, A-law, etc. The detailed treatment of the preceding algorithms is beyond our scope.

You know that file extension of any file uniquely identifies the file type. This is true for audio files also. However, from the audio file extension (for example .wav file) it is not always possible to know the exact type of the CODEC. Take, for example, the .wav extension. The *.wav* files are the standard audio format of Microsoft and IBM. The files with .wav extension always represent audio files. However, a .wav file may be created using a variety of CODEC algorithm—they may be ADPCM encoded, µ-law encoded, A-law encoded or even uncompressed. How will the CODEC decompressing the compressed file know which CODEC was used to compress the file? The answer is, by reading the header information in the file. Every audio file (except a raw file) has a header that specifies exactly which format is being used.

The only formats that do not employ compression are raw audio files. Also, it does not have any header information. So to read from a .raw file, you have to prompt the sampling frequency (441. Khz, 22.05 Khz, etc), the sample size (16-bit and 8-bit, etc.) and the number of channels.

Depending on the operating system of your computer and the CODEC supported, you may find any one or more of the following audio CODECs (for Windows-9x PCs, select: My Computer → Properties → Device Manager → Sound Video and Game Controller → Audio codecs → (double click) → Audio Codec properties (aee Table 3.1).

Table 3.1 *Some Standard Audio CODECs*

CODEC or file type	File Extension	Remarks
International **M**ultimedia **A**ssociation version of ADPCM IMA-ADPCM CODEC	*.wav*	A good alternative to MPEG with fast decoding and good quality of compressed audio.
Microsoft Microsoft ADPCM	*.wav*	Uses ADPCM; not as efficient as IMA-ADPCM CODEC.
CCITT standard G.711 standard formats A-law or µ-law *.wav*	*.wav*	8-bit per sample files created from 16-bit per sample using A-law or µ-law encoding
Raw files Raw sample values with no header	*.raw*	When a raw file is opened, you are required to give the sampling rate, resolution and number of channels.
NeXT/Sun (Java)	*.au*	Can be both uncompressed or compressed using either of existing techniques (such as CCITT µ -law, A-law, etc.), mono or stereo, 16 or 8 -bit and different sampling rates when uncompressed; often used in Java applications and applets and distribution on the Internet.
MPEG-1 Layer 3 audio	*.mp3*	High compression rate with good quality; presently the leading web-based format.
Advanced Audio Coding	*.aac*	Designed as an improvement over *.mp3* files; this lossy CODEC is used by iPods and cell phones.
Quick Time Movie	*.mov* (sometimes *.qt*)	A multimedia container file framework created by Apple; used for video, sound, and text in separate tracks. Cross-platform - for Mac, PC and Linux. Uses a variety of audio and video codecs, including Sorensen, MPEG, Cinepak, and DivX.

This list is not exhaustive.

When applying lossy compression to audio or video data in a multimedia project due to limitation of storage and streaming, it is always advisable to work with a copy and to keep the uncompressed original digital audio data backed up so that you may use it in future. Also the issue of using the CODEC that will be available across different operating system platforms is vital.

Lossless Compression Techniques

Lossless compression algorithms are applied when you cannot afford to loss data. Such lossless algorithms are routinely applied in commercial compression utilities, such as *pkzip*, *winzip*, *compress* (on UNIX), etc., for compression of text and image files. Audio files normally do not yield good results with lossless compression, as repetition of same signal consecutively is not typical. Lossless algorithms are sometimes used as one of the steps in a hybrid algorithm, such as the JPEG compression algorithm that also employs lossy algorithm. Here, a few of the important lossless algorithms that are mainly used for text, image and binary encoded computer program compression are discussed.

Run Length Encoding

Run Length Encoding (RLE) is one of the simplest forms of lossless data compression. It is mainly used in compression of image, text and binary encoded

NOTES

Run-Length Encoding (RLE): One of the simplest forms of lossless data compression

computer programs where loss of data cannot be afforded. In principle, if an information contains symbols (for digital image—the color values of consecutive pixels or for text—the consecutive characters in a text file) that have the tendency to repeat continuously, then instead of coding each symbol in the group individually, RLE stores the symbol code and the number of times it is repeated in the group, thus saving space. RLE is not generally applied to audio files because you can very rarely find a piece of audio where the sound intensity does not change over time.

A Practical Example of RLE: The Microsoft version of bitmap image files (the *.bmp* files) has the option to use run length encoding. An image file stores the series of color values for consecutive pixels across rows and columns. You know that if an image is stored in the RGB color mode, each pixel uses three bytes—one for each of red, green and blue channels. For ease of understanding, let us consider an image in grayscale, i.e., each pixel encoded in 1 byte (you can subsequently extend the concept to 3 bytes for r, g and b channels). One byte can represent 2^8 or 256 gray scales. So a grayscale image file will store a sequence of numbers from 0 to 255 representing the grayscale values of pixels across the rows from left to right for each row and then subsequent rows down the column.

Now, if you consider a grayscale image of size 640 × 480 pixels, you will know that the image will store 640 × 480 = 3,07,200 pixel information in as many bytes. In the previous image information, there may be some grayscale values that are repeated quite frequently. The idea behind RLE is that instead of storing each of the 3,07,200 pixels as an individual value, it will store pixels information as number pairs (c, n), where c is a specific grayscale value that is repeated over n consecutive pixels. For example, let the first 10 pixels in the previous 640 × 480 grayscale image file are:

0 0 0 118 118 118 242 0 255 255

the RLE of this sequence will be

(0, 3), (118, 3), (242, 1), (0,1), (255, 2)

Intuitively, you can figure out that the more a symbol is repeated continuously, the more compression can be achieved by the RLE method.

Entropy Encoding

Entropy compression makes statistical analysis of the frequency of symbols to achieve compression by assigning one code word to each symbol and at the same time assigning shorter code words to encode more frequently occurring symbols. Examples of entropy compression are Shannon-Fano, Huffman encoding and adaptive Huffman encoding.

Claude Shannon, the famous engineer and mathematician, was the first to throw light on the limits of lossless compression and suggest methods for achieving better compression using entropy encoding. He introduced the concept of *entropy of a discrete random event* S (where S can have possible states 1, 2, 3 …. n) as :

$$\eta = H(S) = \sum_{i=1}^{n} p_i \log_2 \left(\frac{1}{p_i} \right)$$

where,

p_i = probability of the i^{th} state happening in the random event S

The preceding equation is known as *Shannon's Entropy Equation*. Shannon borrowed the term entropy from physics. You may wonder why Shannon used the term entropy. Entropy generally means chaos or disorder. In other words, it indicates the various ways in which an event can happen. For instance, if you roll a pair of dice, you can get five in four different ways, whereas you can get two (or twelve) in only one possible way. So, in this context you can say that five has greater entropy than two (or twelve).

The Shannon's entropy equation tells us the optimum value of the average number of bits required to represent each instance of a symbol in a string of symbols *S*, depending upon the frequency of each symbol appearing in *S*. Shannon proved that the optimum value of the average number of bits required cannot be further reduced, i.e., no further compression can be achieved.

Let us take an example:

If you consider the information source *S* as a grayscale image where each pixel can randomly take any grayscale value from 0 to 255 (i.e., 2^8 or 256 different values), then the probability of a pixel having each grayscale value is 1/256. The Shannon's equation then gives the optimum value of the average number of bits as $\sum_{0}^{255}\frac{1}{256}\left(\log_2(256)\right)=\sum_{0}^{255}\frac{1}{256}(8)=8$ bits (or 1 byte). Note that since you use 1 byte per pixel to represent grayscale values in an image, it is already optimized as per Shannon's equation, and you cannot further compress it.

Let us take another example:

This time let the information source S be the string of characters in the word: 'shannon'. So the word has 7 characters with the following frequencies in Table 3.2.

Table 3.2 *Frequencies of Individual Characters in the Word 'Shannon'*

Character	**Frequency**	**Optimum number of bits required to encode this character [A]**	**Relative frequency of the character in the word [B]**	**Product of columns [A] and [B]**
S	1	2.807	0.143	0.401
H	1	2.807	0.143	0.401
A	1	2.807	0.143	0.401
N	3	1.222	0.429	0.524
O	1	2.807	0.143	0.401

Then Shannon's equation becomes,

$$\eta=\frac{1}{7}\log_2\left(\frac{7}{1}\right)+\frac{1}{7}\log_2\left(\frac{7}{1}\right)+\frac{1}{7}\log_2\left(\frac{7}{1}\right)+\frac{3}{7}\log_2\left(\frac{7}{3}\right)+\frac{1}{7}\log_2\left(\frac{7}{1}\right)=$$
$$0.401+0.401+0.401+0.524+0.401=2.128$$

This means that the minimum average number of bits required to encode each character in the word 'Shannon' is at least 2.128.

Shannon-Fano Algorithm

You have understood that the Shannon's equation implies that to achieve a better compression ratio, you need to use different number of bits to represent different

symbols depending on their frequency of occurrence in an information source *S*. Fewer bits are required to represent symbols that appear more frequently. This is also known as *Variable Length Coding (VLC)*.

The *Shannon-Fano algorithm* is a variable length coding method used for lossless compression of data that applies the concept of minimum average number of bits required to encode symbols in an information source *S* as per Shannon's equation.

The algorithm takes a recursive top-down approach, dividing the symbols into two parts, such that the symbols in the two halves have approximately the same frequencies of occurrence in the file being compressed. The algorithm generates a code-tree where the branches are labelled with 0s and 1s; the string of 0s and 1s obtained down each path from root to the leaf nodes produces the code for the symbol related at the leaf node.

You will understand the scheme through the following example.
Let us text a text string (the same word 'Shannon', which you will used to explain the Shannon's equation). The encoding steps are as follows:

(i) Symbols are sorted on the descending frequency of occurrence in the string.

So, the characters of the word *shannon* is sorted as n, s, h, a, o and the frequency Table 3.3 is created.

Table 3.3 *Frequency Table for the Word 'Shannon' sorted on Descending Frequency*

Character	**Frequency**
N	3
S	1
H	1
A	1
O	1

(ii) Next the symbols are divided into two parts—each with approximately same number of frequencies till all leaf nodes contain only one symbol.

In the example, when you divide the symbols (i.e., the characters) into two groups N (count = 3) and (S, H, A and O total count = 4) the binary tree will look like the following: The frequencies of occurrence are shown beside the characters in Figure 3.2.

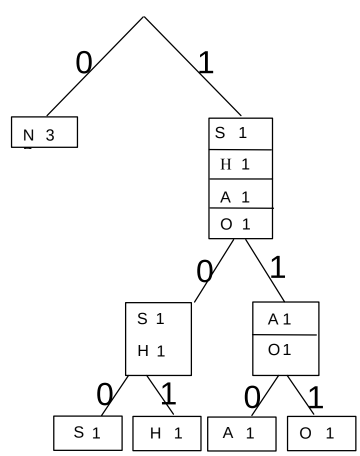


Fig. 3.2 *Example of the Shannon-Fano Algorithm as applied to Compression of the Word 'Shannon'*

Based on the Figure 3.2, you can now list the codes assigned to each character as shown in Table 3.4.

Table 3.4 *List of Codes assigned to Each Character*

Character	Frequency [1]	Optimum number of bits required to encode this character [2]	Code [3]	Number of bits used per character [4]	Total Number of bits required [1] x [4]
N	3	1.222	0	1	3
S	1	2.807	100	3	3
H	1	2.807	101	3	3
A	1	2.807	110	3	3
O	1	2.807	111	3	3
				Total number of bits required	15

The average number of bits used as per Table 3.4 to represent each character in the word *shannon* is 15/7 = 2.143. As per Shannon's equation, the minimum number is 2.128. Hence, you are very close to the lower bound.

Huffman Encoding

Huffman encoding is another lossless compression employing entropy encoding which is more efficient that the Shannon-Fano algorithm in terms of compression rate.

NOTES

Huffman encoding: Another lossless compression employing entropy encoding

It is normally used on text and bitmap image files. In this algorithm also, a variable length-encoding scheme is used such that the characters (for a text file) or the colors (for an image file) that appear more frequently are encoded with fewer bits. So you can term it as a form of entropy encoding, such as the Shannon-Fano algorithm, etc., described earlier.

The Huffman encoding algorithm performs the job of compression in the following steps:

It determines the codes (of variable lengths) for each of the characters or colors in a text or image file respectively. The information is stored in a frequency table containing individual character/color and its frequency of occurrence. A tree data structure is developed based on the frequency table. Unlike the Shannon-Fano algorithm, which employs a top-down approach, the encoding scheme in Huffman algorithm takes a bottom-up approach. This is explained in the subsequent section.

From the frequency Table 3.5 select two symbols (text-character or color) having the lowest frequency of occurrence as two nodes. These two nodes are joined to form a parent node. The frequencies of the children nodes are added and assigned to the parent node. A Huffman sub-tree is thus created. Insert the parent node in the list maintaining the ascending order. Delete the children from the list and continue this cycle till the list has only one symbol left.

During the preceding process, any two nodes (having the least frequencies in the list) that do not already have a parent node can be combined. Hence, sometimes you have to combine two leaf nodes, sometimes two parent nodes or sometimes a leaf node and a parent node.

To illustrate the Huffman algorithm, let us consider the same word 'shannon' and proceed to create a frequency table and a binary tree.

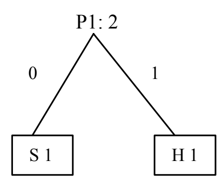
NOTES

Step 1: Form the frequency Table 3.5.

Table 3.5 *Frequency Table for the Word 'Shannon' sorted on ascending Frequency*

Character	Frequency
S	1
H	1
A	1
O	1
N	3

Step 2: Form and expand the Huffman sub-tree and modify the list created from the frequency Table 3.5.

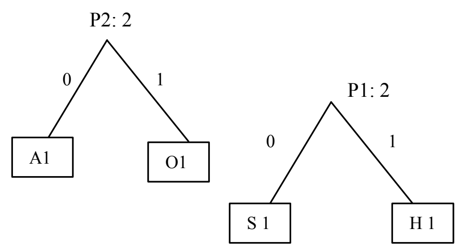


The list sorted in ascending order with frequencies in bracket:

S(1) H(1) A(1) O(1) N(3)

After creation of the Parent node P1 insert it in the list and delete the child nodes.

~~S(1) H(1)~~ A(1) O(1) P1(2) N(3)



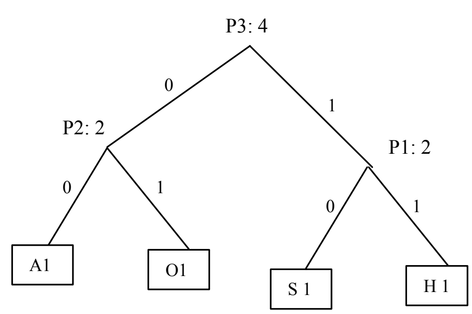
The revised list (deleting S and H):

A(1) O(1) P1(2) N(3)

After creation of the Parent node P2 insert it in the list and delete the child nodes.

~~A(1) O(1)~~ P1(2) P2(2) N(3)

Note that now there are two Huffman sub-trees created has and they have to be linked together.



The revised list (deleting A and O):

P2(1) P2(2) N(3)

After creation of the Parent node P3 insert it in the list and delete the child nodes.

~~O(1) P2(3)~~ N(3) P3(4)

Step 3: Based on the preceding Huffman-tree, you can now list the codes assigned to each character, as shown in Table 3.6.

NOTES

Table 3.6 Codes assigned to each Character

Character	Frequency [1]	Optimum number of bits required to encode this character [2]	Code [3]	Number of bits used per character [4]	Total Number of bits required [1] x [4]
S	1	2.807	110	3	3
H	1	2.807	111	3	3
A	1	2.807	100	3	3
O	1	2.807	101	3	3
N	3	1.222	0	1	3
				Total number of bits required	15

The average number of bits used as per Table 3.6 using the Huffman encoding algorithm to represent each character in the word *shannon* is 15/7 = 2.243. This is the same value obtained from the Shannon-Fano algorithm. Note that as per the Shannon's equation the minimum number is 2.128. Thus, you are close to the lower bound.

Having obtained the codes for the symbols, next, the text/image file is compressed by replacing each character/color with its code.

Observe that in the preceding example, the symbols *S*, *H*, *A*, and *O* all have the same frequency of occurrence, i.e., 1. If two or more minimum value nodes have the same value, the choice of which one to use is arbitrary. Thus, there may be more than one Huffman tree, and more than one possible way an information source (text or image file) may be encoded. You could have started with *A* and *O* instead of *S* and *H* and obtained a different set of code words. However, the average number of bits will be the same.

A unique feature of the Huffman coding is that no code is a prefix of another code. For example, you will not find a code 0 *and* a code 00, or 000. Similarly, you will not get a code 1 along with the codes 11 or 111 or 110. This is because the codes are created from the tree data structure. Otherwise, it would have created tremendous problem since the codes are of variable length.

Arithmetic Encoding

One disadvantage with the Shannon-Fano and Huffman algorithms is that each symbol in the information source has to be treated separately. Accordingly, each symbol is given a unique code, represented by an integral number of bits.

However, from Shannon's equation, you have seen that one can achieve optimum encoding by using a non-integer number of bits for each code. For example, you have already learned that in order to represent the text string 'Shannon' if you could use exactly 2.128 bits to each character, you could achieve an optimum compression rate. However, with the earlier two algorithms, you can only use integer number of bits. *Arithmetic encoding* overcomes some of the drawbacks of the Huffman and Shannon-Fano algorithms by encoding a file or group of symbols as an entity instead of assigning a code to each symbol. Here, a group of symbols is encoded using a single floating point number. Otherwise, the arithmetic encoding is also based on statistical analysis of the frequency of symbols in a file.
Arithmetic encoding uses the same strategy as entropy encoding:

Step 1: Begin with a list of the symbols in the information source (a text or image file) and their frequency of occurrence.

NOTES

You can keep the list unsorted or you can sort them in ascending or descending order of frequencies, but whatever you choose, the same list should also be used while decoding.

Step-2: Next, express the frequencies as numbers between 0 and 1 and assign each symbol a probability interval.

To start with, let us define f_{low} = Low value of the frequency = 0

f_{high} = High value of the frequency = 1

$p_{interval}$ = Probability interval = $f_{high} - f_{low}$ = 1-0 = 1

s_{klow} = The lower value in the probability interval for the k^{th} symbol

s_{khigh} = The higher value in the probability interval for the k^{th} symbol

Step-3: Select the k^{th} symbol in the string (k = 1,2,3,….) of symbols to be coded (a character in a text string or the color of a pixel in a string of pixels) and calculate the code sub intervals as follows:

$$f_{low\text{-}new} = f_{low} + p_{interval} * s_{klow}$$

$$f_{high\text{-}new} = f_{low} + p_{interval} * s_{khigh}$$

Step-4: Update the code sub intervals. Then select the next symbol.

$$f_{low} = f_{low\text{-}new}$$

$$f_{high} = f_{high\text{-}new}$$

$$p_{interval} = f_{high} - f_{low}$$

Repeat Steps 3 and 4 till the last symbol is processed.

Step-5: On coming out of the loop, store the number f_{low} as the encoded value

This is illustrated with the help of an example.

Consider a text string containing the characters 'O, S, C, H, L' as per Table 3.7. Let the first six letters in the preceding text string be – 'SCHOOL' and apply the arithmetic encoding on it.

First, you have to create the list of symbols (the characters and their frequency)— the characters (symbols) taken in arbitrary order:

Table 3.7 *Lists of Symbols and the Frequency of their Occurrence*

Character	Occurrence	Frequency	Probability interval	
			s_{klow}	s_{khigh}
O	20	20/100 = 0.2	0	0.2
S	30	30/100 = 0.3	0.2	0.5
C	10	10/100 = 0.1	0.5	0.6
H	10	10/100 = 0.1	0.6	0.7
L	30	30/100 = 0.3	0.7	1
Sum	**100**	**1**		

You have already accomplished Steps 1 and 2. Notice that the sum of all the frequencies is 1 and the frequencies are expressed as real numbers between 0 and 1 and each character is assigned a frequency interval whose size matches with the frequency of occurrence. For example, the frequency interval of the character 'S' is 0.2 to 0.5.

Next, you have to go through the Steps 3 and 4 recursively to narrow down the probability interval by calculating the code sub-intervals as per Table 3.8.

Table 3.8 *Calculations of the Code Sub-Intervals*

character	f_{high}	f_{low}	$p_{interval}$	$f_{low\text{-}new}$	$F_{high\text{-}new}$
S	1	0	1	0+1*0.2 = 0.2	0+1*0.5 = 0.5
C	0.5	0.2	0.3	0.2+0.3*0.5 = 0.35	0.2+0.3*0.6 = 0.38
H	0.38	0.35	0.03	0.35+0.03*0.6 = 0.368	0.35+0.03*0.7 = 0.371
O	0.371	0.368	0.003	0.368+0.003*0 = 0.368	0.368+0.003*0.2 = 0.3686
O	0.3686	0.368	0.0006	0.368+0.0006*0 = 0.368	0.368+0.0006*0.2 = 0.36812
L	0.36812	0.368	0.00012	0.368+0.00012*0.7 = 0.368084	0.368+0.00012*1 = 0.36812

NOTES

From the final step, the low value obtained (i.e., 0.368084) gives the encoded value of the entire string 'SCHOOL' in the text file.

The reverse process does decoding. For instance, to get back the string 'SCHOOL' from the encoded value 0.368084, you can take the following steps:

From the initial probability table (See Table 3.8) the number 0.368084 lies within the probability range assigned to 'S' (0.2 to 0.5). So 'S' is the first symbol in the string.

Now to go back one step above the code sub-interval subtract the low value (s_{klow}) of the symbol 'S' from the encoded value and divide it by the probability interval ($p_{interval}$) of 'S', this results in:

(0.368084 – 0.2)/(0.5–0.2) = 0.56028. This value lies in the probability interval range of 'C'. So 'C' is the second symbol in the string.

Now to go further back by one step above the code sub-interval subtract the low value (s_{klow}) of the symbol 'C' from the encoded value and divide it by the probability interval ($p_{interval}$) of 'C', this results in:

(0.56028 – 0.5)/(0.6–0.5) = 0.6028. This value lies in the probability interval range of 'H' (0.6 to 0.7). So 'H' is the third symbol in the string.

Similarly, all the other symbols are extracted/decoded.

Though the arithmetic encoding/decoding process is workable for the previous example, you will appreciate that to represent the string of symbols (characters in a text string or pixel information is an image) by a real number you will require very high precision floating point arithmetic, which may not be practical. In reality, actual arithmetic encoding algorithms employ bit shifting operations and integer arithmetic, thus avoiding requirement for any floating point operations at all. IBM and other companies have filed a number of international patents on such sophisticated encoding techniques and many legal battles have been fought over their use.

3.6 IMAGE COMPRESSION STANDARDS

Image compression denotes compression of data for digital images. The objective of such compression is the reduction of redundancy of data pertaining to the image data that makes storage and transmission of data in an efficient form. But such reduction of data may lead to loss of some data pertaining to the image and for this reason, an image compression may prove to be loss.

Image compression: It denotes compression of data for digital images

Thus, there are two types of image compression, lossy and lossless. Lossless compression is preferred in technical drawing, medical imaging, comics and icons. A lossy compression, when applied at low bit rate, is subject to compression artifacts.

NOTES

3.6.1 Methods Used in Image Compression

The following methods are generally used in image compression

Lossless Compression: In this method data is compressed in such a way that on decompression the display would be an exact replica of the original data.

- **Run-length Encoding:** This is default method used in PCX and one of the methods used in BMP, TIFF and TGA formats.
- **Entropy Encoding**
- **DPCM and Predictive Coding**
- **Adaptive Dictionary Algorithms:** LZW is an algorithm used in TIFF and GIF image formats.
- **Deflation:** This method is used in PNG, TIFF and MNG formats.
- **Chain Codes**

Lossy Compression

- **Reduction of Color Space:** In lossy compression, color space is reduced to some of the most common colors of the image. Colors, so selected, are specified in the header of compressed images' color palette. Every pixel makes a reference to the index of a color in this color palette. To avoid posterization, this method is combined with dithering.
- **Chroma Subsampling:** This method makes use of the fact that human eye has perception for spatial changes more in case of brightness in comparison to that of color. This averages or drops some of the information on chrominance of the image.
- **Transform Coding**: This method is most commonly used and in it a transform that is Fourier-related is applied. This is followed by quantization and entropy coding.
- **Fractal Compression:** This technique works on the principle of self similarity.

Fractal compression: This technique works on the principle of self similarity

The main objective of image compression is the production of best quality of image at a given bit-rate, also known as compression rate.

3.6.2 JPEG Image Compression Standard

In multimedia, different signals for digitization are either time-varying (audio) or space-varying (still image) or vary with both time and space (video). The number of times a time-varying periodic signal (such as audio, radio or light) changes during a unit of time is the frequency of the signal. Similarly, for a space-varying signal such as the color intensities of pixels in an image, frequency or more specifically spatial frequency is the number of times the intensities varies over a unit of distance. The component frequencies (for different colors) and amplitudes (color values) for each frequency can be computed for an image using Fourier Transform and can be displayed graphically in the frequency domain. The horizontal axis represents frequency and the vertical axis represents

amplitude in such frequency spectrum graph—the spikes at different frequencies correspond to different signal components (see Figure 3.3).

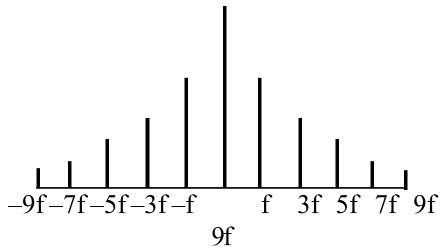


***Fig. 3.3** Frequency Spectrum*

The spike at zero frequency is called the DC component, which can be considers as the average value of the signal. The other (non-zero) spikes are called AC components. The high-frequency components are associated with abrupt and frequent changes in pixel intensity across the image. Psychophysical experiments suggest that people do not perceive the effect of high frequencies (wide variation of color within small area) very accurately. So, they are less likely to notice any change if some of the very high special-frequency components are removed from an image-signal. JPEG's approach is basically to remove special redundancy that is to reduce high-frequency components of an image and then the result is coded into a bitstring.

The JPEG encoder (shown in Figure 3.4) works in the following manner:

Step 1: Transform RGB to YIQ or YUV and subsample color.

Step 2: Perform DCT on image blocks.

Step 3: Apply quantization.

Step 4: Perform zigzag ordering and run-length encoding.

Step 5: Perform entropy coding.

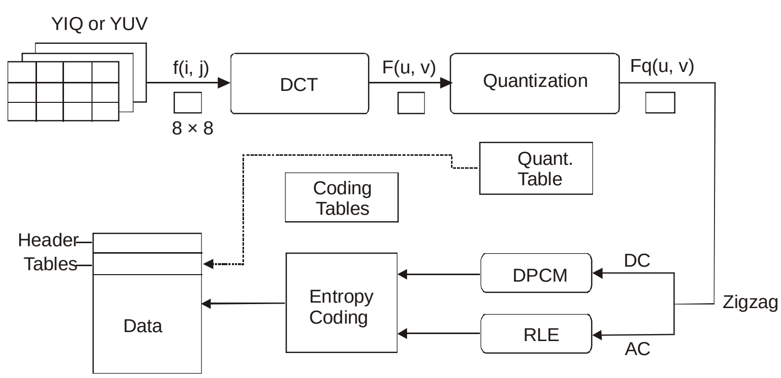


***Fig. 3.4** Block Diagram for JPEG Encoder*

Step 1: JPEG works for both color and gray scale images. In case of color images in YIQ or YUV format, the JPEG encoder works on each component separately using the same routines. If the source image is in a different color format, say RGB, the encoder performs a color space conversion to yield YIQ or YUV image signal. When the JPEG image is needed for viewing, the three compressed image components can be decoded independently and eventually combined.

Step 2: An image is a function say $f(i,j)$ of the two dimensions x and y. JPEG uses 2D Discrete Cosine Transform (DCT) to transform the image into its frequency components. In computational terms, DCT takes an array of pixel values $f(i,j)$ and produces a same sized two-dimensional array of coefficients $F(u,v)$, representing the amplitude of the special frequency components in two directions.

It shows that DCT computation takes the form of a nested double loop, iterating over the two dimensions of the array. The computational time for each DCT coefficient is proportional to the image size in pixels ($i \times j$) and the entire DCT computation time is proportional to the square of the size.

Step 3: To reduce the total number of bits needed for a compressed image, DCT coefficients for different frequencies are quantized to different levels with fewer levels being used for higher frequencies. JPEG uses a 64-element **quantization table**, which must be specified by the application (or user) as an input to the encoder.

Step 4: Before entropy coding of compressed data, two additional lossless compression steps are carried out on the quantized DCT coefficients. These are run-Length Encoding **(RLE)** on AC coefficients and Differential Pulse Code Modulation (DPCM) on DC coefficients.

Step 5: The DC and AC coefficients finally undergo entropy coding. Only the basic (or baseline) entropy coding method which uses Huffman coding and supports 8-bit pixels is used. Each DPCM coded DC coefficient is represented by a pair [SIZE, AMPLITUDE] where SIZE indicates how many bits are needed for representing the coefficients and AMPLITUDE contains the actual bits. Huffman coding scheme uses 1 bit for the most frequently occurring SIZE, 2 bits for the next most frequent SIZE and so on saving in space for storing DPCM coded DC coefficient data. Huffman codes are thus Variable-Length Codes (VLC) and the coding requires that one or more sets of Huffman code tables be specified by the application which uses JPEG. The same tables that are used to compress an image are also needed to decompress it.

JPEG Modes

The JPEG standard supports numerous modes (variation). Some of the commonly used ones are as follows:

- Sequential mode
- Progressive mode
- Hierarchical mode
- Lossless mode

  - **Sequential Mode:** The key difference with the sequential mode is that each image component is encoded in multiple scans rather than in a single scan.
  - **Hierarchical Mode:** This mode provides a 'pyramidal' or hierarchical encoding of an image at multiple resolutions, each differing in resolution from its adjacent encoding by a factor of two in either the horizontal or vertical dimension or both.
  - **Lossless Mode:** Lossless JPEG is a very a special case of JPEG, in which, indeed has no loss in its image quality. It does not use DCT-based method. Instead, it employs a simple differential (predictive) coding method.

Sequential mode: The key difference with the sequential mode is that each image component is encoded in multiple scans rather than in a single scan

3.6.3 MPEG Motion Video Compression

MPEG (Moving Picture Experts Group) is the international standard for audio and video digital compression and MPEG-1 is most relevant for video at low data rate (upto 1.5 M bit/s) to be incorporated in multimedia. MPEG-1 is standard with five parts, namely—systems, video and audio, conformance testing and software simulation (a full C-language implementation of the MPEG-1 encoder and decoder). Though higher standards like MPEG-2, MPEG-4, MPEG-7 and MPEG-21 have evolved in search of a higher compression ratio, better video quality, effective communication and technological upgradation, you will study MPEG-1 only for understanding of the basic MPEG scheme.

NOTES

MPEG-1

MPEG-1 standard does not actually explain a compression algorithm but it defines a datastream syntax and a decompressor. The datastream architecture is based on a sequence of frames, each of which contains the data that is needed to create a single displayed image. There are four different kinds of frames (depending on how each image is to be decoded, which are as follows:

1. ***I-Frames*** (intra-coded images): They are self-contained and are coded without any reference to other images. These frames are spatially compressed using a transform-coding method similar to JPEG. The compression ratio for I–frames is the lowest within MPEG. An I–frame must exist at the start of any video stream and also at any random access entry-point in the stream.
2. ***P-Frames*** (predictive-coded images): They are compressed images resulting from the removal of temporal redundancy between successive frames. These frames are coded by a forward predictive-coding method in which the target macroblocks are predicted from most similar reference macroblocks in the preceding I or P–frame. Only the difference between the spatial location of the macroblocks, that is, the motion vector and the difference in the content of the macroblocks are coded. Instead of the difference, a macroblock itself is coded as non-motion compensated macroblock when a good match as reference macroblock is not found. P-frames usually achieve a large compression ratio (three times as much in I–frames).
3. ***B-Frames*** (bi-directionally predictive-coded frames): They are coded by interpolation between two macroblocks—one from forward prediction (from previous I or P-frame) and the other from backward prediction (from future I or P-frame). Interpolative motion compensation is used here. If matching in both directions is successful, two motion vectors will be sent and the two corresponding matching macroblocks will be averaged (interpolated) for comparing to the target macroblock in order to generate the difference macroblock. If an acceptable match can be found in only one of the reference frames, then only one motion vector and its corresponding macroblocks and used for generating the difference macroblock. Maximum compression ratio (one and half times as much as in P-frame) is achieved in B-frames.
4. ***D-Frames*** (DC-coded frames): They are intraframe-coded and are used for fast forward or fast-rewind modes.

D-Frames (DC-coded frames): They are intraframe-coded and are used for fast forward or fast-rewind modes

NOTES

3.6.4 Digital Video Interface Technology

The Digital Display Working Group created the Digital Video Interface (DVI) technology that accommodates for interconnecting the single connector to both digital and analog interfaces. The DVI technology supports high-speed connection (digital) to display the videos and animations. It facilitates a common interface for all the display devices. The high-speed digital signals in DVI technology support up to 350 Mpixels/sec as well as SVGA. The plug and play service is supported via hot-plug detection in this interface. It is frequently being used in LCD displays. The nature of operation of this technology is implemented in the pure digital form. Therefore, intermediate analog signals are not needed to use the DVI technology. It encounters analog-to-digital conversion to remove the synchronization of the operations and aliasing problems. Conceptually, DVI implies a collection of video interface through which LCD flat monitors get good quality modern video-graphics cards. For this, DVI cables are being used including both VGA and DVI output port (see Figure 3.5).



Fig. 3.5 *Digital Video Interface*

The technology behind DVI uses Transmitting Minimized Differential Signaling (TMDS) to transmit the desired data over DVI connection. One TMDS is linked to transmit the data but sometimes dual links and two TMDS channels are also preferred to link if it is assembled to the system unit. The three data channels (RGB fiber optics) are maintained by a single link in which one channel (clock control channel) is used to access the demanded video even if the dual link is connected to the system unit (see Figure 3.6).

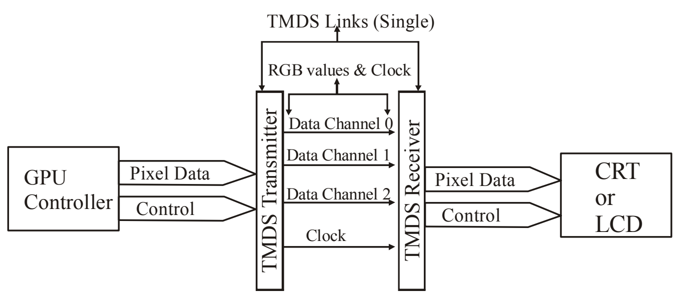


Fig. 3.6 *Data Channels Linked with DVI Connection*

In Figure 3.6, the Graphics Processing Unit (GPU) controller is used to send the pixel data that is passed to various data channels. The data channels are named data channel 0, data channel 1 and data channel 2 that are bound with time constraints. After controlling and checking, each pixel is used in graphics and the specific data are sent to the CRT or LCD monitors to be processed that are to be viewed by the customers/viewers. The TDMS link of 10-bit is generally operated till 165 MHz. It can also be linked up to 1.65 Gbps of bandwidth. A digital flat panel is displayed by

1920×1080 screen resolution and the electric power is generated at 60 Hz. Basically, this flat panel supports dual-link TMDS because dual link supports 2Gbps of bandwidth and is operated to match every second link to the previous one. The dual TMDS uses 2048 × 1536 screen resolution to achieve better graphics for the DVI technology.

NOTES

Working of DVI

The system unit (PC/VGA/CRT) creates and transmits the video signal in the form of digital signals (0 and 1). The digital CRT monitor displays the analog signals but the video card of the system unit (VGA connection) converts the digital signals to analog ones to display the video data. The role of DVI technology is important at this step because the LCD monitor uses the graphics interpreter with the help of the DVI connection and changes analog signals back to digital ones.

In Figure 3.7, most of the video cards are used with DVI technology to convert the digital-to-analog signal and then convert analog-to-digital signal in the LCD display unit. The data in the PC is sent to the processing of electronic signals and then it appears on the LCD monitor.

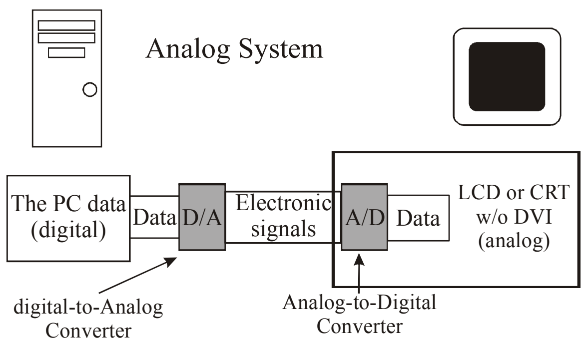


Fig. 3.7 *LCD Monitor with DVI Technology*

The image quality is not good and the display unit shows a lower resolution. The function of DVI is to remove the bad quality of graphics that appear in the video or animation.

Types of DVI

The DVI represents the mode of video interface technology, which is used to optimize Digital Flat Panel (DFP) standards. The types of DVI connections are as follows:

1. DVI-D (True Digital Video)

In the DVI-D system, the cables are directly connected to the source video, for example, video cards and digital LCD monitors. It offers a high quality image in comparison to the analog image. The analog signal is sent to the monitor and then changed to the digital format. It enhances the source connection and omits the process of analog conversion.

2. DVI-A (High-Resolution Analog)

In the DVI-A system, the cables are used to transmit the DVI signal to display in the analog format for the LCDs and CRT monitors.

DVI-A (High-Resolution Analog): In the DVI-A system, the cables are used to transmit the DVI signal to display in the analog format for the LCDs and CRT monitors

NOTES

Check Your Progress

11. Fill in the blanks with appropriate words.
   a. _________ metrics is used as prime tool to compress image and data.
   b. _______ compression denotes compression of data for digital images.
   c. ________ in a digital video image occurs when the same information is transmitted more than once.
   d. ______ is a format for bitmap images, introduced by CompuServe in 1987.
12. State whether the following statements are true or false.
   a. If the compressed data are properly indexed then it improves the performance of mining data in the compressed large database as well.
   b. Redundancy in digital video occurs when the same information is transmitted more than once.
   c. Temporal redundancy is removed by compressing each individual image frame in isolation and the techniques used are generally called spatial compression or intra–frame compression.
   d. MPEG is the international standard for audio/video audio compression.

3. DVI-I (High-Resolution Analog)

In this system, the cables are integrated and are used to transmit the digital source signal to a digital display or it can change the analog source to an analog display. It makes the DVI-I connection flexible and better than other DVI technologies.

Features of DVI Technology

The features of DVI technology are as follows:

1. It uses proprietary chips and the data compression method to create a form of multimedia that is to be integrated into the desktop system unit.
2. It is helpful to play back the full motion videos, multiple stereo sound tracks, live television shows, color graphics, etc.
3. It incorporates and stores DVI into the storage devices for desktop system unit and Winchester hard drives.
4. It plugs the interface boards to distribute the available expression slot on the motherboard and then installs the software.
5. It provides an enhanced technique manifesting the text-oriented e-mail and marks the sender of the message with the motion video communication.
6. It supervises the information tracking to transmit the videographic instructions for subordinating the messages sent, as per VoD technology.
7. It is used to send and receive the applications of videographic presentations, videographic voice mail, audio–visual databases, audio–visual references, sales messages, etc.

3.7 SUMMARY

In this unit, you have learnt that:

- Digital multimedia object files are normally very large. For the purpose of storage and transmission, they are required to be compressed.
- Compression algorithms are divided into two fundamental types (i) Lossless compression and (ii) Lossy compression.
- In a lossless compression, no data or information is lost at the time of compression and decompression process. While compression condenses the size of the file, the decompression process restores the data back in its original value and size.
- In lossy compression sacrifices some information. However, the information that is sacrificed utilizing the limitations of human vision or hearing and the loss of fidelity is not perceptible to a human being.
- Multimedia data compression technique is used to reduce the redundancies in data representation with reference to decrease in the data storage requirements and hence communication overloads when transmitted through a communication network.
- Redundancy in a digital video image occurs when the same information is transmitted more than once.

- Run-length encoding is default method used in PCX and one of the methods used in BMP, TIFF and TGA formats.
- Out of different multimedia elements the need for compression is greatest for video as the data volume for Full Screen Full Motion (FSFM) video is very high.
- Image compression denotes compression of data for digital images. The objective of such compression is the reduction of redundancy of data pertaining to the image data that makes storage and transmission of data in an efficient form.
- Transform coding method is most commonly used and in it a transform that is Fourier-related is applied. This is followed by quantization and entropy coding.
- MPEG (Moving Picture Experts Group) is the international standard for audio/video digital compression.
- In the DVI-A system, the cables are used to transmit the DVI signal to display in the analog format for the LCDs and CRT monitors.
- In DVI-I, the cables are integrated and are used to transmit the digital source signal to a digital display or it can change the analog source to an analog display. It makes the DVI-I connection flexible and better than other DVI technologies.
- DVI technology is helpful to play back the full motion videos, multiple stereo sound tracks, live television shows, color graphics, etc.

3.8 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Compression algorithms are divided into two fundamental types (i) Lossless compression and (ii) Lossy compression.
2. Multimedia data compression technique is used to reduce the redundancies in data representation with reference to decrease in the data storage requirements and hence communication overloads when transmitted through a communication network.
3. Redundancy in a digital video image occurs when the same information is transmitted more than once.
4. When a scene is stationary or only slightly moving, there is redundancy between frames of motion sequence—the contents of consecutive frames in time are similar, or they may be related by a simple translation function. This type of redundancy is called temporal redundancy.
5. In loss less compression, data is compressed in such a way that on decompression the display would be an exact replica of the original data.
6. Run-length encoding is default method used in PCX and one of the methods used in BMP, TIFF and TGA formats.
7. Transform coding method is most commonly used and in it a transform that is Fourier-related is applied. This is followed by quantization and entropy coding.
8. MPEG (Moving Picture Experts Group) is the international standard for audio/video digital compression.

NOTES

9. In the DVI-A system, the cables are used to transmit the DVI signal to display in the analog format for the LCDs and CRT monitors. In DVI-I, the cables are integrated and are used to transmit the digital source signal to a digital display or it can change the analog source to an analog display. It makes the DVI-I connection flexible and better than other DVI technologies.
10. DVI technology is helpful to play back the full motion videos, multiple stereo sound tracks, live television shows, color graphics, etc.
11. (a) Error, (b) Image, (c) Redundancy, (d) GIF
12. (a) True, (b) True, (c) False, (d) False

3.9 QUESTIONS AND EXERCISES

Short-Answer Questions

1. When is multimedia data compression technique used?
2. When redundancy in digital video does occurs?
3. What are two types of redundancy?
4. List various methods used in image compression.
5. Write short note on working of DVI.

Long-Answer Questions

1. How can we calculate MSE and PSNR? Explain
2. Explain JPEG image compression standard.
3. What are different types of JPEG modes? Discuss with the help of examples.
4. Explain different kind of frames decoding in MPEG-1 standard.
5. What are different types of DVI? Describe each type.

UNIT 4 OBJECT ORIENTED MULTIMEDIA

NOTES

Structure

4.0 INTRODUCTION

Digital multimedia technology has evolved tremendously over the last two decades. Multimedia architecture in terms of hardware and software resources have improved with the improvement of processing power, network bandwidth, data compression technology and cross platform multimedia data transfer protocols and standards. With the improvement of digital multimedia technology, the user interfaces have now become multimedia enabled, offering much superior flexibility in terms of improved and user-friendly interface to software applications. Newer hardware components, such as Universal Serial Bus (USB), IEEE1394 and Small Computer System Interface (SCSI) as well as data transmission protocols, such as Bluetooth, etc., have made multimedia data capture, transmission and playback faster and easier. The most noticeable step towards multimedia-enabled applications is the shift from the text based-user interface to the present day Graphical User Interface (GUI). In this unit, you will learn about the multimedia-enabled applications with reference to the GUI. You will also study about

NOTES

multimedia network, multimedia database architecture and distributed multimedia computing over networks.

The various multimedia elements, such as audio, video, image, and text are used to interface with software applications. Operating systems, such as Windows, McIntosh, X-Window, etc., include support for such elements. Normally, the GUI elements are not embedded into the operating system kernel, but they are available as a set of Application Programming Interface (API), which are hardware-level instructions or low-level functions for implementing various multimedia-based features. The software collection that provides the GUI by accessing the API is called the implementation of the API. It usually contains the documentation and tools for using the API in the software, and is collectively called Software Development Kit (SDK). In Windows, the API is a set of functions written in C and C++ and implemented as Dynamic Link Libraries (DLL) mainly in the operating system core files. Examples of such APIs include Direct3D, DirectX and OpenGL.

Large repositories of digital audio and video data are available commercially in multimedia database over the Internet. The famous *YouTube* is an example, which offers thousands of video clippings over the Internet on diverse subjects. The other examples are the availability of videoconferencing facility and video-on-demand (set-top boxes) at affordable price. This has become possible due to standardized streaming technologies and real-time multimedia playback protocols, such as RTP and RTSP and standardized playbacks architecture, such as the Windows media framework and QuickTime framework. An important matter related to the distributed multimedia application is effective synchronization between parallel playing data streams like video and audio.

Finally, this unit will focus on object oriented approach and multimedia documents. Objects are a collection of attributes which directly represent structural and behavioral knowledge of a domain. Thus, an attribute is a mapping from a set of objects to a set of objects. When the attribute returns a set of one element (known as singleton set), it is viewed as returning an object rather than a set. SGML is an ISO-standard (ISO 8879:1986) technology. It is used to define generalized markup languages for documents. Markup describes the structure and other features of a document. The ODA enabled documents may contain text; geometric graphics information in Computer Graphics Metafile (CGM) format, or bit-mapped, raster or facsimile or other graphics information. MHEG defines standards of information coding and is defined in ISO/IEC 13522. Subsequently, various revisions have been done to keep up with developments in multimedia.

4.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Understand user interfaces
- Describe hardware support
- Explain streaming technologies
- Understand the concept of MMDBS
- Discuss the object-oriented approach
- Describe multimedia documents, such as SGML, ODA and MHEG

4.2 USER INTERFACES

NOTES

The user interface or Human Computer Interface (HCI) is the collective means by which a human being interacts with a computer system including a peripheral device, a computer program or the computer itself. Normally, the user interface offers a process of *input* by which the user can influence and control a system; and the *output* whereby the system indicates the effect of the manipulation. For instance, in the earlier days of the mainframe computers, an input to the computer system was given mainly by typing commands on the console or by punch cards and the output was obtained at the display terminals and on paper.

Early computer operating systems and applications used what is known as command line interface. Here, text was the only medium for interaction with the terminals as the user was restricted to give commands through the keyboard in a single line of text on the screen, called the command line. This mode of interface is also known as text-based interface. Even before that, in the primitive days of computers, the user had to enter a range of addresses through register switches to give commands and supply data.

Now, we have multimedia user interfaces or computer interfaces that communicate with a user using multiple media objects, such as image icons, text, speech, etc.

4.2.1 Graphical User Interfaces

Graphical User Interfaces (GUIs) are the most popular type of computer interfaces today. They can be used intuitively and are much easier to use than the command-line interfaces. The user can interact with on-screen simulations of familiar objects that give idea about the function of the application they represent. For instance, the icon of a calculator indicates a calculating program or a recycle bin (or a trash can) indicates the folder containing the deleted files.

Graphical User Interfaces (GUIs): The most popular type of computer interfaces today. They can be used intuitively and are much easier to use than the command-line interfaces

The main features of a GUI interface are—on-screen desktop, display windows, options menu, command icons, dialog boxes and online help. The above features are briefly discussed as follows:

The *on-screen desktop* is the screen you normally see in your PC— it emulates your working table in real life. The various graphic elements, such as application icons, buttons, links, dialog boxes and sub-windows are displayed on the desktop.

The main feature of a GUI is the *display window.* It is a rectangular area of the screen used to display a program or various types of output including multimedia data. There is a horizontal bar called the title bar at the top of each window that displays the name of the application and the file accessed in the display window. Multiple display windows may be opened for running multiple programs and applications in a multitasking environment. The display window also features a scroll bar at the side or bottom to scroll through a large document.

An *option menu,* as the name suggests provides a set of options. Users can select options they want by highlighting the option and clicking on it with the mouse, as in the case of selection of the text font or font size in the MS Word program.

NOTES

The *command icons* GUIs are icons representing common actions, such as opening, saving or printing files. These icons may be displayed in a row near the top of the screen called a toolbar. When the mouse pointer is positioned on an icon, the screen tips a one or two-word identification label is displayed. Clicking the icon launches the associated action.

A *dialog box* is a window that appears temporarily for the user to input specific information at run time. It disappears once the user enters the requested information. GUIs routinely use dialog boxes to prompt user responses and provide information. Interactions between the user and software are carried out through common GUI elements, such as text boxes, check boxes, tabs, option and buttons, etc.

The GUI interface also offers *online help* feature. Clicking the help button causes a dialog box to appear asking the user to specify the kind of help needed. The program then searches from either the client machine or online documentation and displays a menu of topics from which a user can make a selection.

It should be kept in mind that the GUIs were made possible with the introduction of mouse technology and multimedia high-resolution monitors. The other variants of the mouse are light pen, track ball and the touch screen. Today, almost all PCs are equipped with a preinstalled GUI operating system, and most application software is designed to work smoothly with them. Xerox PARC as the interface for the Xerox Alto computer introduced the concept of GUI. Apple Inc. made GUI popular with its Macintosh line of computers in early 1980s. At present Microsoft Windows is one of the most widely used GUI operating systems. The GUI consists of graphical window gadgets (WIDGETS), such as windows, menus, check boxes, radio buttons, and icons and utilizes a pointing device, such as a mouse, light pen or touch screen over and above the keyboard. The user can use the pointing device, such as mouse, touch screen, light pen, etc., to interact with both the textual and graphical objects by clicking, moving, dragging over the objects. The GUI is sometimes also referred to as WIMPs that is **W**indows, **I**cons, **M**ouse and **P**ointer. GUI interfaces have become an essential component of multimedia applications. The other operative systems, such as UNIX also offer GUI interface with the X-Windows system. Also, programming languages, such as Java and C have also adopted GUI interface through external APIs though they were originally textual in nature.

Today GUI has evolved as a separate discipline that involves technological issues, such as the use of multimedia hardware elements to use text, image and audio based commands as well as understanding of human cognition to make the interface meaningful and memorable and easy to learn. Also, the GUI should be implementable without too much cost involvement.

4.2.2 Widget Toolkit

Widget toolkit is a collection of widgets, often implemented as a library, for a specific user-computer interaction. The widget toolkits are used for designing applications with GUIs. A typical widget toolkit contains the graphical interface element, such as the text box, check box, button, radio buttons, icons, menu, window, toolbars, scroll bars, etc.—using which a user interacts with the computer. Some of the widgets in a widget toolkit helps in interaction with the user, such as the check boxes, buttons, etc, while some widgets function as containers that contain a group of widgets attached to them, such as windows and panels.

The widget toolkit itself is software with an API that is generally provided with an OS (Operating System) or Window Manager. The widgets in a widget toolkit should adhere to a uniform look and feel (design specification) so that the user in general feels a sense of consistency among various portions of the application, as well as various applications within a GUI.

NOTES

The GUI for a program may be constructed by adding widgets on the top of existing widgets in a cascading manner. For instance, the desktop is itself a widget, over which several widgets (such as toolbars, etc.) may be added or removed. In many implementations, separate application windows may be added on the desktop by the Window Manager, each window being associated with a particular application containing a group of widgets (such as tool bar, scroll bars etc.), which can be viewed and accessed by the applications. Often the low-level widgets are integrated with the OS and interact directly with the OS while the high-level widgets come as separate application program. When the widget is activated (by the click of a mouse on a radio button, for example), an event is detected, and it is passed onto the application.

Among those integrated within the OS are the windows API for Microsoft Windows, the Mac OS toolbox and Apple Macintosh. Examples of high-level widget toolkits for UNIX are GTK+ and Motif used in desktop environment for the X-Window system. Microsoft uses the Microsoft Foundation Classes (MFC) for its own programs and also Windows Forms (which are .NET classes) for handling GUI controls. Qt is another widget toolkit that is available across different platforms like Windows, Macintosh and UNIX platforms. Cross platform toolkits for Java programming includes the Standard Window Toolkit (SWT) and Abstract Window Toolkit (AWT) and more recently Swing from Sun Microsystems.

GTK+

GTK+ is a widget toolkit used for constructing GUI interfaces. It is one of the most popular widget toolkits for the X-Window System, along with Qt. It is used in the Gnome Desktop GUI as the widget toolkit and forms the base of the Gnome desktop. The important features of GTK+ are its flexibility to change the look and feel of the GUI, the ability to render smooth anti-aliased graphics, support for object oriented programming support, extensive support of Unicode character sets (it supports international characters using UTF-8), elegant text rendering and layout using *Pango* and accessibility ATK. The flexibility of GTK+ allows GNOME applications to be ported on other OS platforms, such as Windows and Mac OS X. Also, The GTK+ library can be used by many programming languages like C++, C, Java, Perl, Python, PHP, etc., which has made it very popular as a cross platform widget toolkit. GTK+ is free software and is licensed under the Lesser General Public License (LGPL) as a part of the GNU Project. The current version of GTK+ is GTK+2, which is however not compatible with the earlier version.

Qt

Qt is a cross platform application development framework. It is used both as a widget toolkit for GUI program development, and also for non-GUI programs, such as console tools, etc. Qt was developed by a Norwegian company Trolltech, and was subsequently acquisitioned by Nokia in 2008. Qt uses an extended version of C++ but allows binding with PHP, Java, Python, etc. Qt is available as free, open source software

Qt: A cross platform application development framework

NOTES

distributed under the GNU Lesser General Public License (LGPL). Apart from being a widget toolkit, Qt also supports non-GUI features, such as APIs for file handling across different platforms, SQL database access, XML parsing, etc., and multithreaded applications. Five different varieties of Qt are available for various platforms as follows:

- Qt for Linux/X11 for X Windows.
- Qt for Mac OS X for Apple Macintosh OS X.
- Qt for Windows for Microsoft Windows.
- Qt for embedded LINUX for embedded Linux in mobile equipments, such as PDAs.
- Qt for Windows free edition (a free version for Windows).

X-Window System

X-Window system: A standard widget toolkit and network protocol to build GUI capabilities on UNIX based networked computers

The **X-Window system** (or X) is a standard widget toolkit and network protocol to build GUI capabilities on UNIX based networked computers. Primarily, it is a protocol and definition of graphics primitives. It does not dictate the styles of the GUI elements (widgets), such as tool bars, windows, buttons, etc., but let the individual client programs handle this. As a result, the look and feel of X-based environments differ widely and different programs using X present drastically different interfaces. In UNIX, the GUI is not built as the part of the OS kernel, so the X is built as an additional application layer on top of the OS kernel. It provides basic capabilities, such as working with desktop windows and interacting via the mouse pointer. It was initially developed as part of Project Athena. The X.org Foundation maintains the open source version of X-Windows. The X Window system in its present form offers a standard toolkit and protocol stack for building GUI on UNIX or most UNIX-like operating systems. Desktop environments, such as GNOME, KDE and CDE use the X Window System.

Another important aspect of X is that it is specifically designed to function in client-server model over network connections. However, in this client-server model adopted by X, the nomenclature is reversed, as the user machine is the server while the applications running are the clients. This may be a bit confusing. However, remember that X looks from the perspective of the application. Since it provides display and I/O services to the application, it takes the role of the server, and the application that uses these services and is thus clients.

The design of X implies the clients and server to work separately, thus increasing the overhead and decreasing the performance. The current protocol version of X is *X11* that was released in 1987.

Motif

Motif is a GUI guideline as well as a widget toolkit (the Xm or motif widgets) for building GUI under the X-Window system. Motif also includes the documentation called motif style guide that tells how a motif user interface should look and behave to be *motif compliant*. It is also an industry standard known as IEEE 1295. It was created by the open software foundation that has now become the open group. Its current version 2.1 provides support for Unicode and is widely used in several multilingual applications. It is distinguished by its three-dimension look for various widgets or user interface elements, such as text boxes, menus, buttons, sliders, etc.

Many consider it to be obsolete in comparison to GTK+ and Qt, which is true to some extent. In fact, Sun Microsystems, a major motif user has switched over to GTK+.

NOTES

4.3 HARDWARE SUPPORT

The following provides the hardware support to multimedia architecture:

4.3.1 Universal Serial Bus (USB)

The Universal Serial Bus (USB) was designed as a better substitute for the serial and parallel I/O buses used in earlier computers. It is not that modern computers no longer come with the earlier versions of serial and parallel ports, but the USB has almost replaced them by providing a much faster and user-friendly interconnection method. All modern peripheral devices, such as keyboards, mice, modems, printers, scanners and even CD-ROM drives, Webcams, digital cameras, iPods, etc., are routinely corrected in the USB. For some devices, such as Webcam, digital camera or scanner USB has been the standard connection. Even some devices charge the batteries through the USB cable.

For a USB device, when the host computer powers up, or when a device is connected to the USB, it searches for all of the devices connected to the bus and assigns each one an address. This is called enumeration. The host also finds out the type of data transfer speed required for the particular device connected. For instance, the *interrupt mode* is chosen for a device, such as a mouse or a keyboard, for which the data sent is of very little volume. On the other hand, for a device, such as a printer, etc., which receives data in big packets, the *bulk transfer mode* is chosen. The host sends data to the printer in units of 64-byte and verifies. Finally, for streaming devices, such as speakers the *isochronous mode* is chosen where data streams are transmitted between the host and the device in real time without any error checking or correction. The USB has the following features:

- It requires a microprocessor-based controller; hence it is used for PC peripherals. The computer acts as the host.
- Up to 127 devices can be connected to the host, normally by way of USB hubs.
- With USB 2.0, the bus has a maximum data rate of 480 megabits per second (or 60 Mega Bytes/second).
- A USB cable has two wires for power (+5 volts and ground) and a twisted pair of wires to carry data. On the power wires, the computer can supply up to 500 milliamps of power at 5 volts.
- Low-power devices (such as mouse, etc.) can draw their power directly from the bus. High-power devices (such as printers, etc.) normally have their own power supplies The USB hubs can have their own power supplies to provide power to devices connected to the hub.
- USB devices are *hot swappable*, i.e., the devices may be connected or unplugged from the USB port at any time.

NOTES

Currently there are four versions of USB:

(i) USB 1.0 supports a greater rate of 1.5 Mbit per second.

(ii) USB 1.1 supports a greater rate of 12 Mbit per second.

(iii) USB 2.0 supports a maximum data rate of 480 Mbit per second.

(iv) USB 3.0 (introduced by Intel and partners in the year 2008) supports a maximum data rate of 5 Gbit per second.

4.3.2 Small Computer System Interface (SCSI)

The Small Computer System Interface (**SCSI** pronounced skuzzy) is a standard for transferring data between devices and computer. It defines the set of commands, and physical interface protocols. A company called Shugart Technology introduced Shugart Technology System Interface (SASI) in 1979, which was the predecessor of the SCSI interface. The name was changed to SCSI when a number of other companies, such as NCR, Adaptec, etc., decided to adopt SASI. The SCSI specification (SCSI-1) was approved by ANSI in 1986, and thereafter SCSI was developed as an industry wide standard. SCSI is mostly used for connecting hard disks and tape drives. However, it is also used for connecting a broad range of other devices, such as scanners, DVD/CD drives, etc.

SCSI: A standard for transferring data between devices and computer. It defines the set of commands, and physical interface protocols

SCSI is an intelligent interface where every device may be attached to the SCSI bus in a similar manner. Up to 8 or 16 devices can be attached to a single bus. There can be any number of peripheral devices but there should be at least one host. SCSI has a provision for error checking and maintains a buffered interface. SCSI is normally used to communicate between host and a peripheral device.

4.3.3 IEEE 1394 (Firewire)

The IEEE 1394 (Firewire) interface is a serial bus interface standard for isosychronous or streaming data transfer and high-speed communications among computers and audio/visual peripheral devices, such as digital camcorders, etc. The interface is also popularly known as *FireWire*, which is the trade name given by Apple Inc. The other names being *Lynx* (Texas Instruments) and *i.LINK* (Sony). IEEE 1394 is not pluggable, and devices may be added or removed from a powered bus.

The IEEE 1394 has been adopted as the standard connection interface for Audio/Visual (AV) component communication and is controlled by the High-Definition Audio-Video Network Alliance (HANA). The IEEE 1394 interface has replaced SCSI interface in many applications because of its simplified and more adjustable cabling system and lower implementation cost. The high-end digital camcorders have the IEEE 1394 interface as the main data transfer mechanism to the computer. These days many computers including laptops that are specially designed for multimedia design and playback have built-in IEEE 1394 FireWire/i.LINK ports.

Check Your Progress

1. Why are GUIs popular?
2. What is on-screen desktop?
3. What are the functions of the option menu?
4. What does a widget toolkit contain?
5. What is Qt?
6. What is X-Window system?
7. What are the different versions of USB?

4.4 STREAMING TECHNOLOGIES

Streaming technologies allow us to view or listen to media files (video/audio) while these are downloaded in real-time from a computer network. The source material for streaming may be either pre-recorded material or live presentations. By clicking a

NOTES

media link on a Web page, the remote server is accessed and the media file starts downloading as an often slow but continuous stream of small packets of information. Depending on the bandwidth limitations, may be due to heavy Internet traffic or poor network condition, the media data stream may at times pauses momentarily or even breaks up. This is called *true streaming*. There is another kind of streaming called *progressive download*. Here, the media file can be played back only after a considerable portion of the media file has been downloaded to the computer. The viewer may save the streaming media file in the client computer for later viewing.

Streaming technologies consist of many interacting hardware and software components that functions together to create, store and deliver media files over the Web. There are basically three major prevalent streaming technologies. They are *QuickTime, RealMedia and Windows Media technology*. Each of these three streaming technologies has the following three components:

(i) Servers and media file specification.
(ii) Media players or Plug-ins (for example, the Quicktime Player).
(iii) Encoding and creation tools.

The media file specifications for QuickTime (.MOV files), RealMedia (.RM files) and Windows Media (.ASF files) each has its own corresponding streaming server specification to stream files in a chosen format.
There are two approaches that can be taken to reduce the bandwidth and streaming media storage space:

- Reduce the height and width of the display dimensions from the conventional size of at least 640 pixels by 480 pixels and/or reduce the frame rate from 30 frames per second. Most streaming videos use a small display window. Reducing both the display size of the video and the frame rate can drastically reduce the size of the streaming media file. However, reduction in frame rate may cause flickering of the image.
- Reduce the file size through compression. As the file sizes are more and more reduced to accommodate lower bandwidths, loss of quality, especially for video becomes more noticeable. The type of compression used for creating streaming media files is generally lossy compression.

There is a limit to reducing the streaming video display size and the frame rate. Excessive reduction will make the video unacceptable. So, compression is almost always applied to streaming media files (which is lossy in nature) sacrificing the quality of the raw video and audio files.

For effective streaming, the raw audio/video files (in Qt, AVI, WAV or AIF formats) must be compressed below the target bandwidth capacity. Otherwise, the presentation may stall. Most audio/video editing programs (such as the RealProducer from Real Media) has provisions for encoding multiple streaming audio/video clips for different target bandwidth (for example, 22.8, 56, 112 Kbps, and so on). The different versions of the compressed files (audio and video clips) are then stored in dedicated streaming servers (such as the Helix Streaming Server) and the streaming software intelligently selects the version based on the available bandwidth.

4.4.1 Streaming Servers

After the media file is created (digitized from the raw audio/video form), compressed and encoded as streaming media files, they are stored in and delivered from a *streaming server*.

A streaming server is actually a server machine connected to a network. It has a set of specialized software for managing the process of delivery of media files over the Internet. Streaming servers are usually more complex in terms of operational management than a conventional Web server. Although a standard web server may be used to host streaming media files, the performance rapidly deteriorates if the streaming media has to be multicast or delivered to a large numbers of viewers. Also, streaming servers are indispensable for synchronized multimedia shows with lengthy media files.

It should be borne in mind that even the most powerful streaming server has limitations in terms of delivering audio/video streams over the network. The number of streams possible to be delivered simultaneously may be a few thousands, but if the limit is reached, the server will be shown as busy.

4.4.2 Streaming Audio Video Formats

The various formats of streaming audio video are discussed in the following sections:

Advanced Systems Format (ASF)

It is part of the Windows Media Framework and Microsoft's Microsoft's proprietary audio video format for streaming. The CODECs offers choice to select different quality settings by selecting either Constant Bit Rate (CBR) or Variable Bit Rate (VBR), and lossy or lossless compression and uses the .ASF file extension.

The Advanced Systems Format (ASF) is a container format that contains the common file types, such as the Window Media Audio (WMA) and Windows Media Video (WMV). ASF files can also contain additional information, such as the title of the album, names of artists, etc., known as metadata.

Flash Video Format (FLV)

Flash video file format: Another very popular container format designed to deliver streaming audio and video over the Internet using Adobe Flash Player

The **flash video file format** is another very popular container format designed to deliver streaming audio and video over the Internet using Adobe Flash Player. To encode FLV files any of the standard tools, such as Adobe Flash, Sorenson Squeeze or On2Flix, etc., are used. This ease of encoding and distributing Flash video has made it extremely popular. A Flash video may be displayed on a web page in either of the following ways:

- By embedding the video within an SWF file and then playing with a Flash Player in a Web page.
- By using progressive download (via HTTP) that allows random-access at any point in the video file.
- Streaming video by means of Real Time Multimedia Protocol (RTMP) from the user's own Flash Media Server or a hosted server using Flash video streaming services.

Ogg Format

The Ogg format is a free open standard audio/video and metadata container framework. Software patents do not cover the different CODECs available. The Xiph Foundation maintains the format. As an open-source format, the Ogg format (file extension .ogg) is good for Internet streaming.

NOTES

4.4.3 MPEG-4 Format

MPEG is an evolving standard, widely accepted by the industry and divided in many parts covering different multimedia element formats, like such as video, audio, subtitle, advance video coding, etc. Out of the different parts, the MPEG-4 Part 2 was designed as a versatile standard, which among other applications, addresses streaming video on the Web at low-data-rate. However, it also addresses high-data-rate HDTV broadcast and DVD playback. Its initial design was based on the QuickTime container format. The MPEG-4 Part 10 standard also known as MPEG-4 Advanced Video Coding (AVC) was introduced in 2003 and is equivalent to H.264. It has achieved wide adoption for CD, DVD, HD-DVD, Blu-ray disc distribution as well as Web streaming media and videoconferencing and broadcast television. MPEG-4 is a versatile and diverse range of formats supporting data rates ranging from 5 Kbits/sec to 10 Mbits/sec.

4.5 MULTIMEDIA DATABASE SYSTEMS (MMDBS)

In this section, you will be introduced to the principles and techniques involved in the design of the basic system of a multimedia database along with the problems and issues involved in the design and implementation of a MMDBS. Remember that MMDBS is still very much an evolving subject in terms of technology and standardization.

You must be familiar that multimedia generally means an integrated set of two or more multimedia objects in digital form, such as text, image, audio, video, graphics and animation. Over the last decade, technological advances in key areas, such as processor technology, affordable high bandwidth over communication networks as well as secondary storage devices, and newer I/O devices have resulted in computer applications that generate and transmit multimedia data at unprecedented scale. MMDBS that combine time independent (text, image, graphics, etc.) and time dependent data have become an integral element of modern information infrastructure. The requirements for storage and retrieval of huge amount of multimedia data have made it essential to design MMDBS that will provide a unified frameworks for storing, processing, retrieving, presenting and transmitting various types of digital media object data, each having different formats. It should be understood that the multimedia objects, such as an image or a video clip cannot be retrieved from the MMDBS just by matching a sample object with another. In a traditional database, queries are made by a keyword, an exact index or by specifying a range; however, in case of MMDBS, the data is inexact and subjective in nature. Hence, such keyword-based or index-based searches become ineffective. For example, if a Website publishes results using a traditional database, you may retrieve the result by giving the roll number. However, the retrieval of records of a student by specifying some facial features from a database of facial

NOTES

images is non-exact and requires content-based or similarity-based queries. So, the problems concerned with the design of a multimedia database system are numerous and quite complex in nature and the MMDBS architecture is much more complicated than traditional text database.

4.5.1 MMDBS Applications

An efficient MMDBS is required for efficient management of the huge amounts of both spatial and temporal multimedia data for its effective use in many application areas, such as:

- Digital libraries.
- Collaborative work on CAD/CAM.
- Online documentation.
- Image, video and audio repositories.
- E-learning portals.
- Art and entertainment.
- Advertisement, retailing and marketing, etc.

4.5.2 MMDBS Architecture

Unlike a text database where the alphanumeric textual data are compared—the characteristics of the multimedia data objects are compared (instead of the multimedia objects themselves being compared). Thus, the MMDBS basically manages two different types of information pertaining to the actual digital multimedia data. They are:

Media data: These are the actual or physical data representing the media objects, such as text, images, audio, video, etc

(i) Media data: These are the actual or physical data representing the media objects, such as text, images, audio, video, etc., that are captured, digitized, processed, compressed and stored.

(ii) Metadata: The metadata or the data about the above media data. The metadata includes the following:

(a) *Media format data* or the information pertaining to the format of the media data, such as the sampling rate, resolution, frame rate, encoding scheme, etc. The media format data is mainly used for presentation of the retrieved data.

(b) *Media keyword data* is the content descriptive data or the keyword descriptions—normally relating to the generation of the media data. For example, for an image, this may include the date, time, and camera model, shutter speed, etc.

(c) *Media feature data* is the content dependent data or the features derived from the media object data. The feature characterizes the media content and is useful for data retrieval. For example, the feature data may contain the texture, distribution of colors, different shapes present in an image, etc. The media keyword data and media feature data are used as indices for querying purpose.

The *metadata* depicts the subject, structure, semantics, etc., of the multimedia data. Appropriate metadata should be available in the MMDBS for the multimedia data so that effective querying and processing can be done. For the metadata to properly

model the multimedia data, domain specific information should be captured to the extent possible. Further, the metadata should be able to depict the data independent of the medium of representation. In practice, an object-relational database system may be adopted to manage the metadata. This provides flexibility with respect to querying the metadata itself. For a multimedia database, metadata may come from two sources—content dependant metadata may be extracted from the media objects using feature extracting programs, SQL triggers or operating system scripts. While the domain-independent metadata, such as description of an image or a keyword are generally associated with the media element, such as a video clip or an image and inserted manually into the database.

4.5.3 Features of MMDBS

The following are the features of MMDBS:

- A MMDBS should have proper environment for using and managing digital multimedia database information. In other words, it should support the various multimedia data types, such as text, image, graphics, audio, video and animation.
- It should also support the traditional DBMS functions, such as database definitions, creation, data retrieval, indexing, views, security, integrity, concurrency, backup and recovery, design, documentation and update/query facilities.
- The MMDBS must support large objects. Such objects may either reside in the main database or the main database may contain the metadata only, and the physical media object database is stored externally with a pointer to the file object used to retrieve the media objects.
- The delivery of temporal data like audio and video must maintain a steady minimum rate. The MMDBS must support isosychronous data transfer.
- MMDBS should support similarity based data retrieval. For this special indexing methods should be available.
- It should provide device independent interface.
- It should provide simultaneous access
- It should provide format independent interface.
- It should provide long transaction of media data.

4.5.4 Multimedia Database Queries

In a relational database, retrieval of data stored is usually done by applying queries. The queries contain predicates that have to be satisfied by any data that is retrieved. For a traditional text, Relational Database Management System (RDBMS), the predicates usually involve partial or exact matching, and value ranges, such as find all students who have scored between 45 and 65. However, the issue gets complex when it comes to a multimedia data query.

The simple way to query multimedia data is to define metadata—keywords associated with the multimedia objects that are entered when the data was entered, which is also known as manual indexing. Generally, all data is classified using the same terminology and a standardized keyword dictionary is maintained.

NOTES

For example, if a user wants to search for the 'Taj Mahal', the query checks the keywords of all images stored in the database. Notice that the images themselves are not queried. There exist some problems with the keyword search approach. For example, keyword classification is subjective and the above search, while showing thousands of images of the famous monument may also show images depicting a famous brand of tea! Also, since key wording is done manually, some error may creep in and some data may be wrongly classified. Finally, adding keyword to media elements is never comprehensive; it is a human intensive task. For instance, in a group photograph, you may miss out in tagging one person or in a video clip you may omit the name of a side character; any query with their names will not yield anything. Moreover, it is a very expensive and error prone task, especially for large volume of multimedia data. On the other hand, key wording allows fast retrieval of data and standard-indexing methods may be used since the keywords (text strings) are supported by every Database Management System (DBMS).

A second method called Content-Based Retrieval/Querying (CBR/CBQ) may also be applied in MMDBS queries. The method is still evolving. Unlike the keyword-based queries, CBR is done by audio and image analysis algorithms. So, the better and more intelligent the algorithm, the better is the retrieval of media data from the MMDBS. Analysis of data generally takes place when the data is stored to the MMDBS producing the metadata. The results of the analysis may be multidimensional indexing structures or simply keywords that describe the data. Queries take place on the metadata; however, the data abstraction may be minimized by attempting to describe the data as completely as possible. The generated data can be low level features, such as lines, shapes, colors and textures from which objects could be identified and retrieved through queries. Similarly, audio data can be queried for sounds, word patterns, intonations, etc. As these algorithms become more and more sophisticated, the amount of human intervention in generating the indexing will be minimized.

The image query techniques may also apply for video, as video may be considered as a sequence of images. However, video is a temporal media object, hence it is theoretically possible to query based on specific scenes or activities like someone cycling or a cloud moving in the sky. The idea of analysing video and detecting actions demands sophisticated and powerful CBR algorithms for intelligent query.

In case of CBR providing queries, exact matches will not be possible in most cases. For instance, if you want to match an image of a cloud formation on the sky with another in the MMDBS, then it is most likely that the two cloud formations, the color of the blue sky or the other image properties will ever exactly match. So, the query language will need to be equipped with predicates (called fuzzy predicates) to allow approximate matches to be made. These techniques are still under active research.

4.5.5 Implementation of MMDBS

Multimedia support in *Oracle database 11G* is available as a special feature offering functional support for management of multimedia data types, such as image, audio and video.

Oracle multimedia uses object data types to describe image, audio and video data. The media data components of these objects may be stored either as a Binary Large Objects (BLOBs) or as references to image data residing in external files (BFILEs).

Metadata may be extracted by Java or PL/SQL methods available in the Oracle multimedia objects available to extract image, video or audio data. Various compression and decompression schemes are also available. Video can be streamed on demand from Real Networks Streaming Server or Microsoft Windows Media Services.

NOTES

Query By Image Content (QBIC) is another multimedia database support system. It also supports content-based retrieval of multimedia data. Searching for images in QBIC is based on similarity and thus is quite different from querying in traditional databases. Here the user provides with an initial image and the database retrieves similar images. The user can then further fine tune the search by selecting some images from the images retrieved and further ask the system for more images similar to this selection. This iterative procedure of searching and browsing is termed 'Query By Example' that narrows down the search space.

The QBIC is a typical example of MMDBS using content-based retrieval. The application consists of three logical steps:

(i) *Database population* or loading the images (usually thumbnails) into the MMDBS. The thumbnails may be stored in the hard disk, while the huge image data is stored in a separate server.

(ii) ***Feature calculation*** involves the analysis of color, texture and shape of the images programmatically.

(iii) *Image query* the final step whereby the system retrieves similar images by iteration. The user can refine the search by choosing one of the retrieved images as a new query.

Feature calculation: Involves the analysis of color, texture and shape of the images programmatically

4.6 OBJECT ORIENTED APPROACH

You have learned that the traditional relational database model with the conventional data types is not sufficient to support storage, management and retrieval of large and diverse multimedia objects. To overcome the difficulties, most of the modern multimedia data models proposed are extensions of object oriented models. In the object oriented database approach, new data types (or classes) are defined together with their methods of operation. It permits the extension of existing data types using sub-typing and also allows the modelling of complex relationships between the stored entities. It is ideal for the definition of abstract media types, modelling of complex structured multimedia objects, and operations on media data units. These capabilities allow the usage of object oriented database systems for multimedia applications dealing with media, such as graphics, images and text. Static multimedia documents with complex structures can be modelled without restrictions. However, for time dependent media, the problems of real-time access and appropriate storage techniques still remain. Research activities are going on to overcome these difficulties by extending the scope of Object Oriented Multimedia Database Systems (OO MMDBS) framework to fit into the object oriented framework.

In short, the object-oriented approach allows modelling of application specific data types and classes including their associated operations. This approach offers some support for multimedia but still lacks features, such as supporting time dependent data, user interaction, content-based query and retrieval techniques.

NOTES

The Object Data Management Group (ODMG) model is followed by most of the object oriented database systems. Based on the ODMG model, the Object Query Language (OQL) was defined and adopted by the ODMG.

Object Query Language (OQL) is a query language standard for object oriented databases. The ODMG developed the OQL. It is similar to Sequential Query Language (SQL). However, it is inherently very complex to cater to the complex structured multimedia objects. Due to this complexity OQL is yet to be fully implemented. However, some of the newer query languages like Enterprise JavaBeans (EJB), Quantum Leap (QL) and Java Data Object Query Language (JDOQL) have evolved out of the OQL standard.

The main difference between the OQL from SQL is that OOL supports nesting of objects within objects. Also, not all SQL keywords are supported within OQL. Keywords that are not relevant have been removed from the syntax. Further, mathematical operations can be performed within OQL statements.

4.6.1 Object, Classes and Related System

Object oriented programming does not change the traditional requirements to the Database Management System (DBMS) when object oriented multimedia comes in discussion.

Objects: A collection of attributes which directly represent structural and behavioral knowledge of a domain

Objects are a collection of attributes which directly represent structural and behavioral knowledge of a domain. Thus, an attribute is a mapping from a set of objects to a set of objects. When the attribute returns a set of one element (known as singleton set), it is viewed as returning an object rather than a set. Also, an attribute that always returns a singleton set is called a singleton-valued attribute whereas an attribute that possibly returns a set of objects is called a multiple valued attribute. Attributes can be categorized into two types:

- Enumerated attributes represents structural knowledge such as name.
- Procedural attributes represents behavioral knowledge, such as making medical certificate.

Objects are categorized into instances (which denote individual or factual knowledge) and classes (defining attributes applicable to similar instances).

One important aspects of the object oriented paradigm is the possibility for creating abstract data types. A lot of applications also use databases with objects of variable size. These applications make demands in large, which are based on conventional database technology, including the ability to model very complex data and the ability to evolve without disruptive effects on the current application. Object-oriented multimedia databases address both sources of complexity by including facilities to manage the software-engineering process.

Objects in a multimedia database have different properties and they participate in a number of relationships with other objects. A multimedia database should be capable of retrieving all of the media types that it supports.

A class defines attributes applicable to its instances, like a type in a programming language. To make use of the sequences of records in the multimedia database two kinds of mechanisms are defined: viewers and loaders. A viewer is used to display a particular kind of media. The purpose of a loader is to prepare a media type for

viewing. A specific loader is associated with each type of media. The difference between a loader and a viewer is that a loader will process the compressed media which expands it into a viewable form.

As we know, multimedia includes images, graphics, and text. The class image has attributes, such as header information, frame, and pixel information. The header information describes additional attributes of images. The attribute frame for example contains bulk data in secondary memory. In the similar way, the attribute pixel contains pixel data allocated continuously in main memory. Image, graphics, and text have the same interface for the same manipulation.

Also the object description of the multimedia can be sorted into the entity object and the relation object. Here, the entity object describes a single kind of media which can be figures and images, and a separate special procedure for the media process. The relation object on the other end makes syntactic relationships between multimedia. For example, customer attributes and house frames.

NOTES

4.7 MULTIMEDIA DOCUMENTS

The following are the significant multimedia documents.

4.7.1 Standard Generalized Markup Language (SGML)

Standard Generalized Markup Language (SGML) is an ISO-standard (ISO 8879:1986) technology. It is used to define generalized markup languages for documents. Markup describes the structure and other features of a document. Authors mark up (comment on) their documents by giving information with regard to the structure, presentation and semantic, alongside the content. In visual markup, tags or commands are employed to specify aspects of the appearance of the text, such as fonts and type sizes. In structural markup, tags identify logical elements of a document, such as headings, lists and tables. .

Actually, SGML is used to define a markup language. An example of a HTML document (one of the markup languages) follows:

```
<!DOCTYPE HTML PUBLIC>
<HTML>
   <HEAD>
      <TITLE>My first HTML document</TITLE>
   </HEAD>
   <BODY>
      <P>Hello world!
   </BODY>
</HTML>
```

The preceding HTML document is divided into a header (here, between `<HEAD>` and `</HEAD>`) and a *body* (here, between `<BODY>` and `</BODY>`). The title of the document is in the header, besides other information about the document. The content of the document is in the body. There is just one paragraph in it, markup with `<P>`.

Check Your Progress

8. Distinguish between true streaming and progressive download.
9. How is a Flash video displayed on a web page?
10. List the appli0cations of MMDBS.
11. What is metadata?
12. How is QBIC applied?
13. What is SGML?
14. What are the major features of ODA?
15. What is layout information?
16. What is the function of MHEG?

NOTES

Theoretically, every SGML document has both a logical and a physical structure. Logically, a document is made up of elements, declarations, attributes, character references, comments, and so on. All these are shown in the document by clear markup. Physically, the document is made up of units known as *entities*. A document starts in a document entity. SGML is not only used for conventional document markup, but it can be used for marking up any type of text. Header, paragraphs, footnotes, sections, hypertext links, tables, images, and so on, are the elements in a SGML text. Every element usually describes three parts (i) a start tag, (ii) content, and (iii) an end tag. The name of the element appears in the start tag (written <element-name>) and the end tag (written </element-name>); Elements may have related properties, called *attributes*. The latter may have values (by default, or set by authors or scripts). Attribute/value pairs show before the final ">" of an element's start tag. In the start tag of an element numerous (legal) attribute value pairs, separated by spaces, may appear. They may appear in any order. Numeric or symbolic names included in an SGML document is called character reference. These character references help in referring to rarely used characters, or those that authoring tools make it difficult or impossible to enter. They begin with a '&' sign and end with a semi-colon (;).

Examples of character references are:

'`<`' correspond to s the < sign.

'`>`' correspond to the > sign.

'`"`' correspond to the ' mark.

'`å`' (in decimal) correspond to the letter '`a`' with a small circle above it.

'`И`' (in decimal) correspond to the Cyrillic capital letter '`I`'.

SGML *comments* have the following syntax:

```
<!-- this is a comment -->
<!-- and so is this one,
  which occupies more than one line -->
```

Between the markup declaration open delimiter ('`<!`') and the comment open delimiter ('`--`'), white space is not allowed. However, it is allowed between the comment close delimiter ('`--`') and the markup declaration close delimiter ('>'). To include a series of hyphens ('`--`') within a comment is a common error. Any information that shows between remarks has no particular meaning.

In any markup norm the angle brackets are used as start and end tag delimiters. However, in an SGML text, it is permissible to use other characters, provided an appropriate tangible syntax is defined in the text of the SGML declaration. For instance, an SGML interpreter may be programmed to parse GML markup, wherein the tags are delimited with a left colon and a right full stop, thus, an: e prefix indicates an end tag: :xmp.Hello, world:exmp. As per the reference syntax, upper or lower case is not important in tag names, thus the three tags: (i) `<quote>`, (ii) `<QUOTE>`, and (iii) `<quOtE>` are similar.

In SGML, tags could be substituted with delimiter strings, for example, two equals-signs (==) at the beginning of a line are the 'heading start-tag', and two equals signs (==) after that are the 'heading end-tag'. One characteristic of SGML is the

NOTES

presumptuous empty tagging, such that the empty end tag `</>` in `<ITALICS>this</>` takes its value from the closest preceding full start tag, which, in this example, is `<ITALICS>` (thus, it closes the most recently opened item). The appearance is hence equal to `<ITALICS>this</ITALICS>`. SGML also permits implied markup, various types of tags and many other not obligatory features.

Every SGML parser does not automatically process every SGML text. However, as the system declaration of the processor can be contrasted to the SGML declaration of the text, it was for all time likely to understand if a text was supported by a particular processor or not. Parsing a SGML document that involves traversing the dynamically retrieved entity graph, finding or imply tags and the element structure, and validating those tags against the grammar.

The SGML equivalent, known as Document Type Definition (DTD). It defines only the structure; DTD describes all the texts of a particular type, in terms of the tags that may be used to mark them. SGML without a DTD (for example, simple XML) is a grammar or a language; SGML with a DTD is a Meta language. A separate specification of style layout complementing the DTD is called *style sheet.* For each tag defined in the DTD, a style sheet provides a rule describing the way in which elements with the tag should be laid out. There may be more than one style sheet for a DTD, providing different appearance to the same structure.

Since HTML tags were not enough for the class of web pages developed over time there was a requirement from the web designers to be able to define their own tags. Though SGML has that facility, it is not completely appropriate for use over Internet. Work on making SGML compatible to the Internet led to the development of eXtensible Markup Language (XML) that gives all the facilities of SGML without the overhead (for complicated parsing) forced by SGML. In fact, XML permits Web designers to define their own DTDs for any type of document and Web pages are freed from the limitations of HTML's definition of a document.

4.7.2 Office Document Architecture (ODA)

The Office Document Architecture and Interchange Format was designed to facilitate the presentation, processing and exchange of documents in an open system across a heterogeneous network. The European Computer Manufacturers Association (ECMA) published the ODA standard in 1985 as ECMA-101. Subsequently, ECMA-ODA was adopted by the ISO and the term 'Office' was changed to 'Open' as standard document architecture for the compound documents in the open system and *ODA* is now the acronym for *Open Document Architecture*.

The ODA enabled documents may contain text; geometric graphics information in Computer Graphics Metafile (CGM) format, or bit-mapped, raster or facsimile or other graphics information. The content may include special characters and other information, such as how the content is to be rendered on an output device. There are separate standards in the ODA for character, raster or bit-mapped graphics as well as geometric graphics.

Although character content in an ODA document roughly corresponds to the SGML standard, the two are mutually incompatible. There are two major features of an ODA document:

(i) ODA character content may have embedded control codes that describe how the document is to be formatted and printed uniformly for the sender and the recipient across the network.

(ii) The recipient based on the constraints and rules set by the sender may edit an ODA document.

NOTES

The ODA encoding of a document includes information on the content as well as the structure of the document, along with the information about how it will appear when rendered on a printed page or other output media. For this reason, an ODA document is said to have a logical view and a layout view. In fact, the information content of an ODA document is in three categories:

(i) **Logical information:** It is the relationship of the components of the content, such as sections, chapters, paragraphs, footnotes, etc. It is independent of the page layout.

(ii) **Layout information:** It pertains to the size, positioning, grouping and other image related properties of the content. The layout information is maintained in a hierarchy of components—page set, composite page, basic page, frame and block. Composite pages may contain nested frames and frames contain blocks. Block is at the lowest level (which actually contains the content).

(iii) **The content:** It comprises the alphanumeric characters and geometric shapes. It may also contain the control characters (new line, tab, etc.)

The ODA model also allows defining a generic logical and layout structure. For instance, an ODA compliant word processing software may support a standard template of an ODA document that may be used for reporting purpose by all the departments of a company.

An ODA document belongs to one of three document architecture classes:

- The formatted document class.
- The processable document class.
- The formatted-processable document class.

4.7.3 Multimedia and Hypermedia Experts Group (MHEG)

MHEG is the latest standard related to multimedia presentation. Just as there is a group for multimedia presentation in audio, video and text in an interactive way, known as Motion Picture Expert Group (MPEG), there is another group that describes interactive television services.



MHEG: The latest standard related to multimedia presentation

MHEG defines standards of information coding and is defined in ISO/IEC 13522. Subsequently, various revisions have been done to keep up with developments in multimedia. The latest version is known as MHEG-5, which was created in November 1994.

MHEG model provides a set of standard method; covering other standards, such as still picture format, **Joint Photographic Experts Group (**JPEG) and different standards of MPEG together to produce multimedia presentation. This provides a system independent presentation standard. This group has created standard set of methods for storage, exchange and display of multimedia presentations.

NOTES

Objectives of MHEG

The following are the objectives of MHEG:

- To offer simple, easily implemental framework, using minimum system resources for multimedia applications.
- To define standard format in digital form for presentations which is interactive, and hardware and platform independent.
- To add features, such as extensibility, expandability and customizability by adding code specific to the application. This creates some dependency on platform, but it is desirable to maintain some kind of individualized specialty.

Multimedia presentation packages that are available in the market are proprietary. They are not hardware independent.

Application of MHEG

These days, MHEG is being used at many places. MHEG-5 is used in United Kingdom and New Zealand for interactive digital television. It has been selected to act as middleware in Hong Kong for their digital broadcasting channels that provide interactive services. People want more and more from multimedia applications, and due to this the demand for MHEG standards are growing. MHEG are being used for the following:

- Encyclopaedia on CD-ROM.
- Interactive books and desktop tutors.
- On demand services for news and videos.
- Home shopping under interactive environment.

Structure of MHEG

MHEG uses Abstract Syntax Notation (ASN) to define standard in a structured manner. Abstract Syntax Notation Version 1 (ASN) is used to write text form of MHEG code. This is covered by ISO standard. Figure 4.1 shown the structure of MHEG.

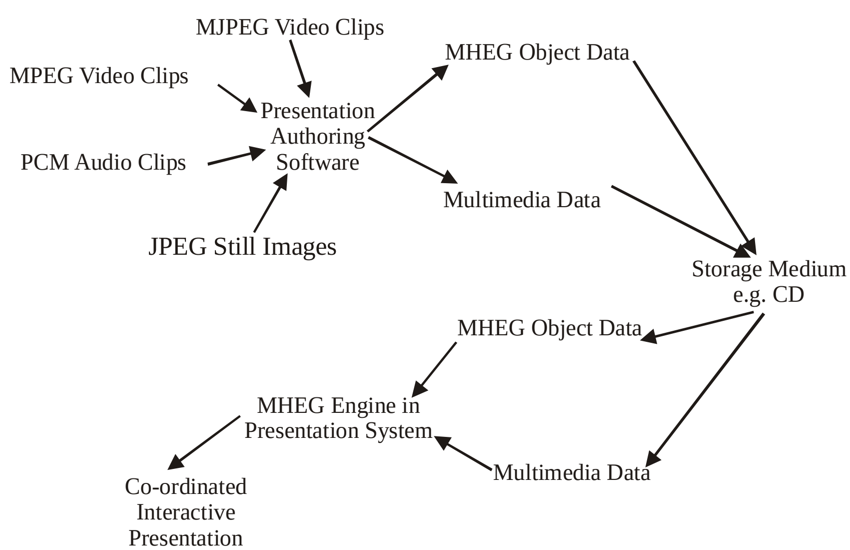


Fig. 4.1 *Structure of MHEG*

MHEG is an object orientated model. It defines many classes. When a presentation is designed, object instances get created. These classes describe the way multimedia components are displayed including interaction with the ongoing presentation. The relationship created among the objects of these classes gives structure for presentation.

Types of Classes in MHEG Model

The following the types of classes in MHEG Models:

- ***Content Classes*:** MHEG is an object oriented, and every component of data in multimedia, such as audio or video has its own MHEG object. If the data is small, such as a small text as title, it is put in the MHEG object. Otherwise the MHEG object provides a reference. This reference may be a filename on a disk.
- ***Behaviour Classes*:** These classes control how and when of data for users. These also permit synchronization of events with user interaction. These classes are of two types (a) action class and (b) link class. They belong to:
  - o Action class when it allows sequential or parallel triggering of events. Example is replay of a number of video clips, one after the other.
  - o Link class when establishing relationships between objects and events. This tells about actions to be taken on objects when a particular event is to be responded.
- ***User Input Classes*:** These are selection and modification classes. These permit a user to select a data or information, provide input to trigger events. Input methods that are defined in MHEG are: Radio button, push button, slider, checkbox, field for text entry, and text lists. These methods enable user to exercise over control of information that is being presented.

Apart from these, there are some other classes also. Some of them deal with the structure of the presentation and object grouping while others deal with interchange of information between machines.

MHEG-5 is popular because it is cost-effective, highly efficient and interactive TV middleware, which has proved itself in the market. It is used two-way communication of TV signals in an interactive way. A wide range of TV centric services are being deployed that enhances viewing pleasure.

Markets

MHEG is being used in the following countries through service providers given under parenthesis:

- In United Kingdom by Freeview (DTT), Freesat (DTH) and TopUpTV (DTT PayTV operator).
- In New Zealand by Freeview (DTH & DTT including HD).
- In Hong Kong - TVB (DTT).
- In India by Digicable (Cable PayTV operator).

NOTES

Content classes: MHEG is an object oriented, and every component of data in multimedia, such as audio or video has its own MHEG object

Check Your Progress

17. Fill in the blanks with appropriate words.
    a. Objects are a collection of __________ which directly represent structural and behavioral knowledge of a domain.
    b. Objects are categorized into instances and________.
    c. Objects in a multimedia database have different _________ and they participate in a number of relationships with other objects.
    d. The header of image class describes additional ________ of images.
18. State whether the following statements are true or false.
    a. When the attribute returns a set of one element it is known as singleton set.
    b. A lot of applications use databases with objects of fix size.
    c. A multimedia database should be capable of retrieving all of the media types that it supports.
    d. The entity object describes a single kind of media which can be figures and images, and a separate special procedure for the media process.

4.8 MULTIMEDIA FRAMEWORKS

NOTES

A multimedia framework is a software framework that handles media on a computer and through a network. A good multimedia framework offers an intuitive API and a modular architecture to easily add support for new audio, video and container formats and transmission protocols. It is specifically designed to be used by applications, such as media players and audio or video editors, but can also be used to build videoconferencing applications, media converters and other multimedia tools.

In contrast to function libraries, a multimedia framework provides a run time environment for the media processing. Ideally such an environment provides execution contexts for the media processing blocks separated from the application using the framework. The separation supports the independent processing of multimedia data in a timely manner. These separate contexts can be implemented as threads.

In identifying abstractions for multimedia programming one should consider the prevailing programming paradigms, such as functional programming, rule-based programming and object oriented programming. The apparent affinity between multimedia and object oriented programming is clearly evident if one looks at the short history of programming environments for multimedia applications.

From the earliest multimedia toolkits, such as Muse and Andrew, to recent commercial multimedia development environments (for example, Apple, Microsoft) one can see the influence of the object oriented paradigm. Often these environments and toolkits, in addition to structuring interfaces into classes and class hierarchies, have the more ambitious goal of building class frameworks for multimedia programming.

Perhaps the main benefits of object oriented technology to multimedia programming are its mechanisms for extending software environments. Many of the issues (media composition techniques, compression schemes, etc.) are at their core, questions of how best to cope with the uncertainties of evolving environments. Frameworks or hierarchies of extensible and interworking classes offer to developers a way of coping with evolution. In the case of multimedia programming, several 'evolutionary processes' are of concern, in particular:

- **Platform Evolution**: The hardware platforms for multimedia applications are rapidly evolving. Capabilities that were once considered exotic, such as video compression and digital signal processing are now found on the desktop.
- **Performance Evolution**: Many of the operations of interest to multimedia programming have real-time constraints, consider audio or video playback as examples. Such temporal dependencies make multimedia applications particularly sensitive to platform performance. It may be necessary, for instance, to adapt to less than optimal processing capacity by reducing presentation 'quality', for example lowering frame rates or sample sizes.
- **Format Evolution**: New data representations for image, audio, video and other media types are likely to appear as a result of on-going standardization activities and research in data compression and media composition.

Developers want to create applications that can adapt to and take advantage of changes in platform functionality, increases in platform performance and new data representations.

Of course, it is impossible to write applications that can fully anticipate future developments in multimedia technology, but frameworks at least offer a mechanism for incorporating these changes into the programming environment.

NOTES

Components of a Multimedia Framework

We now look at a particular multimedia framework which is one that provides explicit support for component-oriented software development. This framework is described more fully elsewhere. In essence it consists of four main class hierarchies: media classes transform classes, format classes and component classes as shown in Figure 4.2.

- **Media classes** correspond to audio, video and the other media types. Instances of these classes are particular media values, i.e., what were called media artifacts.
- **Transform classes** represent media operations in a flexible and extensible manner.

  For example, many image editing programs provide a large number of filter operations with which to transform images. These operations could be represented by methods of an image class; however, this makes the image class overly complicated and adding new filter operations would require modifying this class. These problems are avoided by using separate transform classes to represent filter operations.
- **Format classes** encapsulate information about external representations of media values. Format classes can be defined for both file formats, such as GIF and TIFF, two image file formats and for "stream" formats for instance, CCIR 601 4:2:2, a stream format for uncompressed digital video.

Media
- Text
- Image
  - *BinaryImage*
  - *GrayScaleImage*
  - *ColourImage*
- Graphic
  - 2dGraphic
  - 3dGraphic
- TemporalMedia
  - Audio
    - *RawAudio*
    - CompressedAudio
  - Video
    - *RawVideo*
    - CompressedVideo
  - Animation
    - EventBasedAnimation
    - SceneBasedAnimation
  - Music
    - EventBasedMusic
    - ScoreBasedMusic

Transform
- ImageTransform
- AudioTransform
- VideoTransform

Format
- TextFormat
- ImageFormat
- GraphicFormat
- TemporalMediaFormat
  - AudioFormat
  - VideoFormat
  - AnimationFormat
  - MusicFormat

Component
- Producer
- Consumer
- Transformer

***Fig 4.2** Four Class Hierarchies of a Multimedia Framework*

- **Component classes** represent hardware and software resources that produce, consume and transform media streams. For instance, a CD-DA player is a component that produces a digital audio stream (specifically, stereo 16 bit PCM samples at 44.1 kHz).

  Components are central to the framework for two reasons. First, the framework is adapted to a particular platform by implementing component classes that encapsulate the media processing services found on the platform. Second, applications are constructed by instantiating and connecting components.

4.9 SUMMARY

In this unit, you have learnt that:

- The user interface or human computer interface is the collective means by which a human being interacts with a computer system including a peripheral device, a computer program or the computer itself.
- Early computer operating systems and applications used what is known as command line interface.
- Graphical User Interfaces (GUIs) are the most popular type of computer interfaces today. They can be used intuitively and are much easier to use than the command-line interfaces.
- The main features of a GUI interface are—on-screen desktop, display windows, options menu, command icons, dialog boxes and online help.
- The main feature of a GUI is the display window. It a rectangular area of the screen used to display a program or various types of output including multimedia data.
- An option menu, as the name suggests provides a set of options. Users can select options they want by highlighting the option and clicking on it with the mouse, as in the case of selection of the text font or font size in the MS-Word program.
- A dialog box is a window that appears temporarily for the user to input specific information at run time.
- The GUI interface also offers online help feature. Clicking the help button causes a dialog box to appear asking the user to specify the kind of help needed.
- Widget toolkit is a collection of widgets, often implemented as a library, for a specific user-computer interaction.
- The widget toolkits are used for designing applications with GUIs. A typical widget toolkit contains the graphical interface element, such as the text box, check box, button, radio buttons, icons, menu, window, toolbars, scroll bars, etc.—using which a user interacts with the computer.
- The widget toolkit itself is software with an API that is generally provided with an OS or Window Manager. The widgets in a widget toolkit should adhere to a uniform look and feel (design specification) so that the user in general feels a sense of consistency among various portions of the application, as well as various applications within a GUI.

NOTES

- The GUI for a program may be constructed by adding widgets on the top of existing widgets in a cascading manner.
- Among those integrated within the OS are the windows API for Microsoft Windows, the Mac OS toolbox and Apple Macintosh. Examples of high-level widget toolkits for UNIX are GTK+ and Motif used in desktop environment for the X-Window system. Microsoft uses the Microsoft Foundation Classes (MFC) for its own programs and also Windows Forms (which are .NET classes) for handling GUI controls.
- GTK+ is a widget toolkit used for constructing GUI interfaces. It is one of the most popular widget toolkits for the X-Window System, along with Qt. It is used in the Gnome Desktop GUI as the widget toolkit and forms the base of the Gnome desktop.
- The important features of GTK+ are its flexibility to change the look and feel of the GUI, the ability to render smooth anti-aliased graphics, support for object oriented programming support, extensive support of Unicode character sets (it supports international characters using UTF-8), elegant text rendering and layout using Pango and accessibility ATK.
- Qt is a cross platform application development framework. It is used both as a widget toolkit for GUI program development and also for non-GUI programs, such as console tools, etc.
- Qt was developed by a Norwegian company Trolltech, and was subsequently acquisitioned by Nokia in 2008. Qt uses an extended version of C++ but allows binding with PHP, Java, Python, etc. Qt is available as free, open source software distributed under the GNU Lesser General Public License (LGPL).
- The X-Window system (or X) is a standard widget toolkit and network protocol to build GUI capabilities on UNIX-based networked computers. Primarily, it is a protocol and definition of graphics primitives.
- The system does not dictate the styles of the GUI elements (widgets), such as tool bars, windows, buttons, etc., but let the individual client programs handle this. As a result, the look and feel of X-based environments differ widely and different programs using X present drastically different interfaces.
- Motif is a GUI guideline as well as a widget toolkit (the Xm or motif widgets) for building GUI under the X-Window system. Motif also includes the documentation called motif style guide that tells how a motif user interface should look and behave to be motif compliant. It is also an industry standard known as IEEE 1295.
- The Universal Serial Bus (USB) was designed as a better substitute for the serial and parallel I/O buses used in earlier computers. It is not that modern computers no longer come with the earlier versions of serial and parallel ports, but the USB has almost replaced them by providing a much faster and user-friendly interconnection method.
- All modern peripheral devices, such as keyboards, mice, modems, printers, scanners and even CD-ROM drives, webcams, digital cameras, iPods, etc., are routinely corrected in the USB.

NOTES

- For a USB device, when the host computer powers up, or when a device is connected to the USB, it searches for all of the devices connected to the bus and assigns each one an address. This is called enumeration.
- The Small Computer System interface (SCSI pronounced skuzzy) is a standard for transferring data between devices and computer. It defines the set of commands, and physical interface protocols.
- SCSI is an intelligent interface where every device may be attached to the SCSI bus in a similar manner. Up to 8 or 16 devices can be attached to a single bus. There can be any number of peripheral devices but there should be at least one host. SCSI has a provision for error checking and maintains a buffered interface. SCSI is normally used to communicate between host and a peripheral device.
- The IEEE 1394 (Firewire) interface is a serial bus interface standard for isosychronous or streaming data transfer and high-speed communications among computers and audio/visual peripheral devices, such as digital camcorders, etc.
- Streaming technologies allow us to view or listen to media files (video/audio) while these are downloaded in real-time from a computer network. The source material for streaming may be either pre-recorded material or live presentations.
- Depending on the bandwidth limitations, may be due to heavy Internet traffic or poor network condition, the media data stream may at times pauses momentarily or even breaks up. This is called true streaming.
- Streaming technologies consist of many interacting hardware and software components that functions together to create, store and deliver media files over the Web. There are basically three major prevalent streaming technologies. They are QuickTime, RealMedia, and Windows Media technology.
- After the media file is created (digitized from the raw audio/video form), compressed and encoded as streaming media files, they are stored in and delivered from a streaming server.
- A streaming server is actually a server machine connected to a network. It has a set of specialized software for managing the process of delivery of media files over the Internet.
- Streaming servers are usually more complex in terms of operational management than a conventional Web server.
- Although a standard Web server may be used to host streaming media files, the performance rapidly deteriorates if the streaming media has to be multicast or delivered to a large numbers of viewers.
- It is part of the Windows Media Framework and Microsoft’s Microsoft’s proprietary audio video format for streaming. The CODECs offers choice to select different quality settings by selecting either Constant Bit Rate (CBR) or Variable Bit Rate (VBR), and lossy or lossless compression and uses the .ASF file extension.
- The Advanced Systems Format (ASF) is a container format that contains the common file types, such as the Window Media Audio (WMA) and Windows Media Video (WMV).

NOTES

- The Flash video file format is another very popular container format designed to deliver streaming audio and video over the Internet using Adobe Flash Player. To encode FLV files any of the standard tools, such as Adobe Flash, Sorenson Squeeze or On2Flix, etc., are used.
- The Ogg format is a free open standard audio/video and metadata container framework. Software patents do not cover the different CODECs available. The Xiph Foundation maintains the format. As an open-source format, the Ogg format (file extension .ogg) is good for Internet streaming.
- MPEG is an evolving standard, widely accepted by the industry and divided in many parts covering different multimedia element formats, like such as video, audio, subtitle, advance video coding, etc.
- An efficient MMDBS is required for efficient management of the huge amounts of both spatial and temporal multimedia data for its effective use in many application areas.
- The metadata depicts the subject, structure, semantics, etc., of the multimedia data. Appropriate metadata should be available in the MMDBS for the multimedia data so that effective querying and processing can be done.
- A MMDBS should have proper environment for using and managing digital multimedia database information. In other words, it should support the various multimedia data types, such as text, image, graphics, audio, video and animation.
- In a relational database, retrieval of data stored is usually done by applying queries. The queries contain predicates that have to be satisfied by any data that is retrieved.
- The simple way to query multimedia data is to define metadata—keywords associated with the multimedia objects that are entered when the data was entered, which is also known as manual indexing.
- A second method called Content-Based Retrieval/Querying (CBR/CBQ) may also be applied in MMDBS queries. The method is still evolving.
- The image query techniques may also apply for video, as video may be considered as a sequence of images. However, video is a temporal media object, hence it is theoretically possible to query based on specific scenes or activities like someone cycling or a cloud moving in the sky.
- Multimedia support in Oracle database 11G is available as a special feature offering functional support for management of multimedia data types, such as image, audio and video.
- Oracle multimedia uses object data types to describe image, audio and video data. The media data components of these objects may be stored either as a Binary Large Objects (BLOBs) or as references to image data residing in external files (BFILEs).
- Objects are a collection of attributes which directly represent structural and behavioral knowledge of a domain. Thus, an attribute is a mapping from a set of objects to a set of objects. When the attribute returns a set of one element (known as singleton set), it is viewed as returning an object rather than a set.

- Objects in a multimedia database have different properties and they participate in a number of relationships with other objects. A multimedia database should be capable of retrieving all of the media types that it supports.
- A viewer is used to display a particular kind of media. The purpose of a loader is to prepare a media type for viewing.
- The entity object describes a single kind of media which can be figures and images, and a separate special procedure for the media process.
- The relation object makes syntactic relationships between multimedia.
- Standard Generalized Markup Language (SGML) is an ISO-standard (ISO 8879:1986) technology. It is used to define generalized markup languages for documents. Markup describes the structure and other features of a document.
- The Office Document Architecture and Interchange Format was designed to facilitate the presentation, processing and exchange of documents in an open system across a heterogeneous network.
- The ODA enabled documents may contain text; geometric graphics information in Computer Graphics Metafile (CGM) format, or bit-mapped, raster or facsimile or other graphics information.
- MHEG defines standards of information coding and is defined in ISO/IEC 13522. Subsequently, various revisions have been done to keep up with developments in multimedia.
- MHEG is an object orientated model. It defines many classes. When a presentation is designed, object instances get created. These classes describe the way multimedia components are displayed including interaction with the ongoing presentation. The relationship created among the objects of these classes gives structure for presentation.
- MHEG-5 is popular because it is cost-effective, highly efficient and interactive TV middleware, which has proved itself in the market. It is used two-way communication of TV signals in an interactive way.

4.10 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Graphical User Interfaces (GUIs) are the most popular type of computer interfaces today. They can be used intuitively and are much easier to use than the command-line interfaces. The user can interact with on-screen simulations of familiar objects that give idea about the function of the application they represent. For instance, the icon of a calculator indicates a calculating program or a recycle bin (or a trash can) indicates the folder containing the deleted files.
2. The on-screen desktop is the screen you normally see in your PC. The various graphic elements, such as application icons, buttons, links, dialog boxes and sub-windows are displayed on the desktop.
3. An option menu, as the name suggests, provides a set of options. Users can select options they want by highlighting the option and clicking on it with the mouse, as in the case of selection of the text font or font size in the MS-Word program.

NOTES

4. Widget toolkit is a collection of widgets, often implemented as a library, for a specific user-computer interaction. The widget toolkits are used for designing applications with GUIs. A typical widget toolkit contains the graphical interface element, such as the text box, check box, button, radio buttons, icons, menu, window, toolbars, scroll bars, etc., using which a user interacts with the computer. Some of the widgets in a widget toolkit helps in interaction with the user, such as the check boxes, buttons, etc., while some widgets function as containers that contain a group of widgets attached to them, such as windows and panels.
5. It is a cross-platform application development framework. Qt is used both as a widget toolkit for GUI program development, and also for non-GUI programs, such as console tools, etc.
6. The X-Window system (or X) is a standard widget toolkit and network protocol to build GUI capabilities on UNIX-based networked computers. Primarily, it is a protocol and definition of graphics primitives. It does not dictate the styles of the GUI elements (widgets), such as tool bars, windows, buttons, etc., but let the individual client programs handle this. As a result, the look and feel of X-based environments differ widely and different programs using X present drastically different interfaces.
7. Presently there are four versions of USB.
   (i) USB 1.0 supports a greater rate of 1.5 Mbit per second.
   (ii) USB 1.1 supports a greater rate of 12 Mbit per second.
   (iii) USB 2.0 supports a maximum data rate of 480 Mbit per second.
   (iv) USB 3.0 (introduced by Intel and partners in the year 2008) supports a maximum data rate of 5 Gbit per second.
8. By clicking a media link on a web page, the remote server is accessed and the media file starts downloading as an often slow but continuous stream of small packets of information. Depending on the bandwidth limitations, may be due to heavy Internet traffic or poor network condition, the media data stream may at times pauses momentarily or even breaks up. This is called true streaming.

   There is another kind of streaming called progressive download. Here, the media file can be played back only after a considerable portion of the media file has been downloaded to the computer. The viewer may save the streaming media file in the client computer for later viewing.
9. A Flash video may be displayed on a web page in either of the following methods:
   - By embedding the video within an SWF file and then playing with a Flash Player in a Web page.
   - By using progressive download (via HTTP) that allows random-access at any point in the video file.
   - Streaming video by means of Real-Time Multimedia Protocol (RTMP) from the user's own Flash Media Server or a hosted server using Flash video streaming services.

NOTES

10. An efficient MMDBS is required for efficient management of the huge amounts of both spatial and temporal multimedia data for its effective use in many application areas, such as:
    - Digital libraries.
    - Collaborative work on CAD/CAM.
    - Online documentation.
    - Image, video and audio repositories.
    - E-learning portals.
    - Art and entertainment.
    - Advertisement, retailing and marketing, etc.
11. The metadata depicts the subject, structure, semantics, etc., of the multimedia data. Appropriate metadata should be available in the MMDBS for the multimedia data so that effective querying and processing can be done. For the metadata to properly model the multimedia data, domain specific information should be captured to the extent possible.
12. The QBIC is a typical example of MMDBS using content-based retrieval. The application consists of three logical steps:
    (i) Database population or loading the images (usually thumbnails) into the MMDBS. The thumbnails may be stored in the hard disk, while the huge image data is stored in a separate server.
    (ii) Feature calculation involves the analysis of color, texture and shape of the images programmatically.
    (iii) Image query the final step whereby the system retrieves similar images by iteration. The user can refine the search by choosing one of the retrieved images as a new query.
13. Standard Generalized Markup Language (SGML) is an ISO-standard (ISO 8879:1986) technology. It is used to define generalized markup languages for documents. Markup describes the structure and other features of a document.
14. There are two major features of an ODA document.
    (i) ODA character content may have embedded control codes that describe how the document is to be formatted and printed uniformly for the sender and the recipient across the network.
    (ii) The recipient based on the constraints and rules set by the sender may edit an ODA document.
15. Layout information pertains to the size, positioning, grouping and other image related properties of the content. The layout information is maintained in a hierarchy of components—page set, composite page, basic page, frame and block. Composite pages may contain nested frames and frames contain blocks. Block is at the lowest level (which actually contains the content).
16. MHEG defines standards of information coding and is defined in ISO/IEC 13522.
17. (a) Attributes, (b) Classes, (c) Properties, (d) Attributes
18. (a) True, (b) False, (c) True, (d) False

4.11 QUESTIONS AND EXERCISES

NOTES

Short-Answer Questions

1. What is widget toolkit? What are its uses?
2. Write a note on X-Window system.
3. What is IEEE 1394?
4. What are streaming servers?
5. What are the features of MMDBS?
6. Write a note on object oriented approach in multimedia.
7. What are two types of attributes?
8. What is MHEG? What are its objective and application?

Long-Answer Questions

1. What are user interfaces? Explain the features of GUIs.
2. What are the hardware supports to multimedia architecture? Explain USB.
3. Discuss streaming technologies. What are the different streaming audio videos formats?
4. Explain the MMDBS architecture.
5. 'In a relational database, retrieval of data stored is usually done by applying queries.' Explain with the help of an example.
6. ODMG model is followed by most of the object-oriented database systems. Explain.
7. Discuss the Open Document Architecture (ODA).
8. Discuss the structure and classes of MHEG.

UNIT 5 MULTIMEDIA ENVIRONMENTS

NOTES

Structure

5.0 INTRODUCTION

In this unit, you will learn about typical multimedia environments, such as CD family, media types, organization and applications of media. The compact disc is a thin, round plastic platter which is 12 cms in diameter and approximately one mm thick, with a hole in the center for a spindle. A disk drive is a peripheral device used to store and collect information. It can be removable or fixed, high capacity or low capacity, fast or slow speed, and magnetic or optical. DVD is also an optical disc storage media format. Its main uses are video and data storage. DVDs are of the same dimensions as compact discs (CDs), but store more than six times as much data. An optical drive is a type of storage medium that stores the content in digital form which is written and read by a low intensity laser. The laser reads data from the reflective surface of an optical disc by measuring surface changes in height and depth.

You will also learn about the various types of media and organization of media. Recent advances in compression, storage and communication technologies support the creation of applications that involve storing and retrieving multiple data types, such as text, audio, video, imagery, etc., collectively referred to as multimedia. The Multimedia File System (MMFS) was specifically designed to provide a high performance network interface for storing and retrieving multimedia data. The MMFS maintains association between related files and also helps in storing multimedia data types, such as MIDI files, still images and video animation frames with a universal standard.

Finally, you will learn about applications of multimedia. In the twenty-first century, IT provides many services like airlines, hotel management, Web publishing etc. Some of these services are explained in this unit.

NOTES

5.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Discuss the significance of CD family
- Identify different media types
- Explain organization of media
- Discuss the various applications of multimedia

5.2 THE CD FAMILY

The compact disc is a thin, round plastic platter which is 12 cms in diameter and approximately one mm thick, with a hole in the center for a spindle. A polycarbonate layer of the CD that has the data impressed onto it, is coated with a mirror like metal film (aluminum or gold). That shiny surface is protected by an overcoating of clear plastic and reflects light in a prism-like effect. Side of the CD that reflects light is available for use, while the opposite side is a silk-screen with the disc's identifying label or logo printed on it.

As far as the CD-ROM is concerned, the main feature of the compact disc is the enormous storage capacity (650 MB) such a small slim disc, with a high immunity from damage. In contrast to floppy disks and other conventional secondary storage media, the entire CD data is stored in one spiral track. Thus, the stored information can be easily played back at a continuous data rate making this ideal for audio and video output.

Types of CDs

One major limitation of CD-ROM is that they cannot be used to store data, but only to read data that was stored on them by the manufacturer. However, there are recordable CD's also known as Compact Disc-Recordable (CD-R). Recording on these CDs is expensive even now and it requires a special device called CD-Writer. Another type of CD is CD-RW (Compact Disc-Rewritable) which not only allows data to be written but also allows erasing, thereby making the CDs reusable. Unlike the CD-ROM there are more number of layers in the **CD-R** (Figure 5.1(a), (b)) and even more layers in the **CD-RW**. Some other terminology for CD formats used in the industry are: Audio CD, Photo CD, **Video CD**, **CD-I** (CD Interactive), **CD-ROM/XA** (Extended Architecture), **CD-WO** (Write Once), **CD-MO** (Magneto Optical), etc. All the above types again fall into different sets of standards (or books) namely Red Book, Green Book, Yellow Book and Orange Book. These standards are named after the color of laser lights used in the drives.

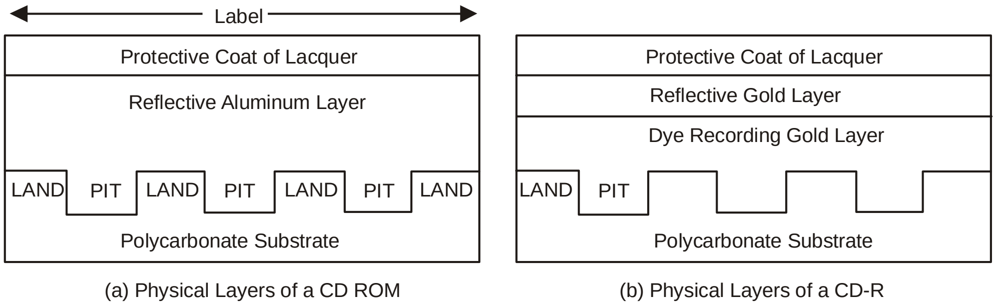


***Fig. 5.1** Difference in the Layers of CD-R and CD-ROM*

1. Recording Process

The CD recording method makes use of optical recording – using a beam of light from a miniature semiconductor laser. Such a beam is of low power, a matter of milliwatts, but the focus of the beam can be to a very small point so that a focussed beam can vaporize the low melting point materials like plastics. Focussing the recording beam to a point on a plastic disc for fraction of a millionth of a second will, therefore, vaporize the material to leave a tiny crater or pit, about 0.6 µm in diameter – for comparison, a human hair is around 50 µm in diameter. The depth of the pits is also very small, of the order of 0.1 µm. The area between these pits, where no beam strikes the disc are called **lands**. Because of the small size of the pits, the tracks of the CD can be much closer – about 60 CD tracks take up the same width as one LP record track.

2. Reading

Reading a set of dimples on a disc also makes use of a semiconductor laser, but of much less power (approximately 780 nm wavelength) since it need not vaporize material. The laser beam after striking the disc will be reflected from the smooth areas (lands) but scattered where there is a pit (see Figure 5.2). By using an optical system that allows the light to travel in both direction to and from the disc surface, it is possible to focus a reflected beam on to a detector, a photodiode and pick up a signal when the beam is reflected from the '**lands**' and with no signal and when the beam falls onto a **pit**. The transition from pit to land and from land to pit corresponds to the coding of a 1 in the data stream. A 0 is called as no transition. Only light from a laser source can fulfil the requirements of being perfectly monochromatic (single frequency) and coherent (no breaks in the wave train) so as to permit focussing on such a fine spot.

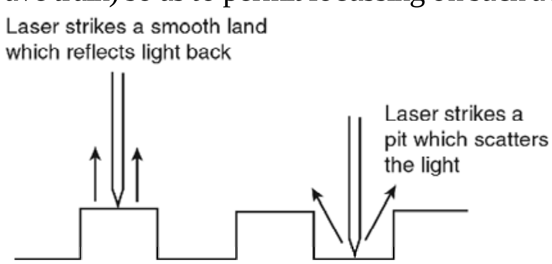


***Fig. 5.2** Reading a Disc*

3. Protecting

The transparent coating over the disc surface focusses the laser beam into the inner layers besides protecting the recorded pits (see Figure. 5.3). Though the

NOTES



Lands: The area between these pits, where no beam strikes the dis

NOTES

diameter of the beam at the pit is around 0.5 µm, the diameter at the surface of the disc (the transparent coating) is about 1 mm. This means that the dust particles and hairs on the surface of the disc have very little effect on the beam, which passes on each side of them – unless the dust particles are a millimetre across. This is just one way in which the CD system establishes its very considerable immunity to dust and scratching on the disc surface the other being the remarkable error detection and correction system, **CIRC** (Cross Interleave Reed-Solomon Code). However the **EFM** (Eight-to-Fourteen Modulation) system used by the CD is itself a considerable safeguard against errors. EFM is basically a modulation system in which a set of eight bits is coded as a set of fourteen bits while recording. The extra bits ensure minimum possibility of error.

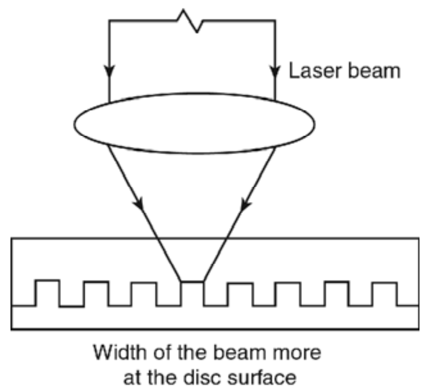


Fig. 5.3 *Protection of a Disc*

4. Quality

The audio performance music CDs is impressive enough with frequency range 20 Hz to 20 kHz (within 0.3 dB) and more than 90 dB dynamic range with the total harmonic distortion (including noise) being less than 0.005 per cent. Added to the audio performance, however, you have the convenience of being able to treat the CD audio (or video) as any other digital signals. The inclusion of control and display data means that the number of items on a recording can be displayed and you can select the order in which they are played and repeating items as per requirement. Even more impressive (especially for music teachers) is the ability to move from track to track, allowing a few notes to be repaired or skipped as required with no risk of damaging the tracks.

5. Speed

You might have noticed the X numbers (52X, 48X) specified with the advertisements of CD drives. These X numbers measure the data extraction rate from the CDs. You should know that the CD drives normally play the CDs at a Constant Linear Velocity (**CLV**) instead of the Constant Angular Velocity (**CAV**) that is used for hard disks. In CLV operation, the disc turns relatively fewer revolutions per second (rps) when it is reading the outer tracks and at a higher (rps) when it reads the inner tracks. What stays constant is the number of bits read each second. A single-speed (1X) CD-ROM drive pumps out data at 150 kbps. Nowadays, 52X CD-ROM drives are mostly available. However, keep in mind that the speed at which you can read data from a CD-ROM is far lower as compared to the reading of data from the hard disk.

CD-ROM Drive

Every CD drive has four major parts:

(a) **The Laser Read Head:** This focusses low power laser for reading a disc and is mounted on a moving arm that enables it to cover the entire disc surface.

(b) **The Motor:** A CD-ROM is held firmly by a spindle system that is connected to a drive motor like the record player. The motor spins the disc faster as the laser read head moves towards the centre of the disc.

(c) **The Prism and Light Sensor:** This prism arrangement channels the laser light returned from the smooth areas on the disc surface back to the light sensor (Photodiode). It moves in tandem with the laser read head.

(d) **The Disc Caddy or Tray:** You place a CD-ROM into the tray and it slides into the drive. The spindle locks the disc into place and it starts spinning.

5.2.1 Introduction to CD Technology

Most of us who work with computers and multimedia tools are familiar with the terms—Compact Disc (CD), CD-ROM, DVD, etc., and routinely use them for retrieval, storage and transportation of digital data. Here, you will learn about the evolution of the CD technology as the most promising secondary storage system and introduce you to the evolution of the different variants of the optical discs and their features. The other optical storage media include holographic data storage and magneto optical data storage devices.

Optical disc refers to any method of storage that uses a laser to retrieve and store data from the media. This term includes such devices as CD-ROM, rewritable optical disc, WORM, CD-R, DVD, Blu-Ray Discs (BD-ROM), etc. The use of optical storage is continuously growing at a fast pace making it a very flexible and affordable medium.

The optical storage technology is constantly evolving in a remarkable pace over the last two decades. Here, you will learn how the optical storage technology has evolved from the CD in 1980s to the present Blu-Ray technology enabling storage of up to 50 GB of data in a single double-layered disc. This overview will introduce you to the types of permanent portable storage, and the background knowledge should help you understand the technology as it continues to evolve.

Way back in the late 1970s and early 1980s, the compact disc was first accepted widely by the music industry as a better alternative to Long Playing (LP) gramophone records. In 1979 electronics giants Sony and Philips together introduced the compact disc technology for delivery of digital audio. Seeing the potential of the compact disc technology for cheap and efficient means to deliver audio, video and other forms of digital data including computer programs other leading companies, such as Microsoft, JVC, etc., joined the collaboration to establish standards for data storage and retrieval.

Optical storage technology has evolved into three basic types of accessing data. They are: Read-Only (ROM), Write Once (R) and Read/Writable (RW). For the Blu-ray, instead of RW the letter E is used signifying the disc is erasable. The abbreviations (CD-ROM, CD-R, CD-RW, DVD-R, DVD-RW BR-E, etc., are printed on the disc and are accepted as industry standard terminology.

NOTES

Optical disc: Refers to any method of storage that uses a laser to retrieve and store data from the media

NOTES

The compact disc technology has evolved in the following stages:

Compact Disc (CD)	• Launched jointly by Philips and Sony in 1979 for delivery of digital music (audio) • Document of standard specifications for audio– **Red Book** • The audio only format specified in the Red Book is also known as CD-DA (compact disk – digital audio) • Manufactured in the factory by stamping the pattern of pits on the spiral track to code the digital audio signals • Read-Only • Predecessor of present optical discs – CD-ROM, DVD etc.
CD-ROM	• Launched by Philips and Sony for storage of digital data after the success of audio CD • Document of standard specifications for data storage and retrieval– **Yellow Book** • Mixed mode of recording (digital data – as per yellow book and audio track – a per Red Book possible) • CD-ROM uses a layer of dye to record data. The process of 'burning' permanently changes the dye forming the *pits*. • Read-Only • Improvement over CD or Compact Disc
CD-RW	• Erasable media – data written on the disc can be erased and re-written • Instead of dye (used in case of CD-R) semi-metal alloy is used. The recording laser melts the alloy and creates the pits – as melting and rapid cooling tarnishes the alloy – making it non-reflective. To erase the disc – the 'pits' are heated by laser of lower power so that on cooling the original crystalline and shiny state comes back • Document of standard specifications for CD-R and CD-RW – **Orange Book** • Improvement over CD-ROM or CD-R
DVD	• Next generation optical storage media • More data stored by decreasing the pit size as well as the pitch (distance between the pits) and also by doubling the layers and sides • Various versions evolved – offering different storage capacities • Document of standard specifications – released by *DVD Forum* – separate books for DVD-R, DVD-RW, DVD-Audio, DVD-Video, etc.
Blue-Ray and HD-DVD	• Next generation DVDs – both used blue laser diodes (shorter wavelength than red laser) to store more data in same disc size • HD-DVD – Developed by Toshiba – abandoned in February 2008 • Blu-Ray – Developed by Sony & Pioneer • Document of standard specifications – released by *Blu-Ray Disc Association* – formed by 9 member companies

As the next-generation high definition optical formats (such as Blu-ray Disc and the now defunct HD DVD) were developed, the original DVD is often termed as Standard Definition-DVD or SD-DVD.

Compact Disc

A Compact Disc (CD), also spelled disk, is an optical disk for the storage of data. CDs were initially built up for the music industry in the late 1970s to store 16-bit, 44.1 kHz digital audio data, holding about 74 minutes of audio track per disk. In the mid-1980s, the first data CDs appeared in market. As already told in the introduction, the audio only format specified in the Red Book is also known as compact disk – digital audio (CD-DA)

$$\text{Audio Data Rate CD-DA} = 16\ \textit{bits/sample} \times 2\ \textit{channels} \times 44100\ \textit{samples/sec}$$
$$= 1{,}411{,}200\ \textit{bits/sec}$$
$$= 176.4\ \textit{Kbytes/sec}$$

The data transferred in each second is stored in 75 sectors. Hence, data stored in each sector is = 1,411,200 / 75 = 18,816 bits = 2352 bytes.

With 74 minutes of playing time for digital audio, capacity of a CD-DA is:

$$= 74\ min \times 1{,}411{,}200\ bits/sec = 6{,}265{,}728{,}000\ bits$$

$$= \frac{6,265,728,000}{\left(8 \times 1024 \times 10^3\right)} = 664\,\text{MB}$$

NOTES

The disc is a thin platter of optical grade polycarbonate, 120 mm in diameter and approximately one mm thick, with a hole in the centre for the spindle. One side of the polycarbonate disc is coated with a very thin mirror like metal film (normally pure aluminium). A coating of lacquer or plastic resin protects the shiny surface so that it does not tarnish due to oxidation by coming in contact with air. Finally, a silk-screened coated label covers the protective coating and bears the title of the album, logo, etc. In the early days when the protective coating technology was not perfect, it often reacted with the aluminium coating and damaged the glossy surface making the CD unusable. To avoid this, gold was often used, as it is an inert metal. However, being too costly, its use has now become very limited, as the protective layer technology has improved. (see Figure 5.4)

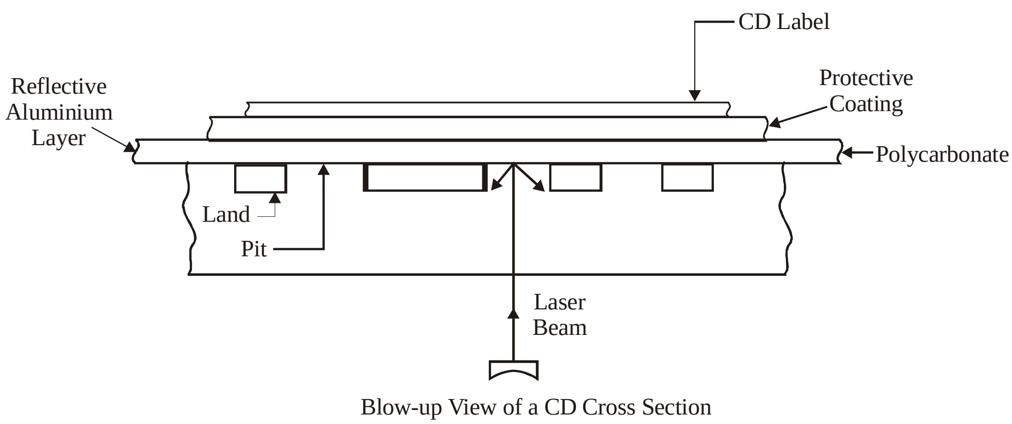


***Fig. 5.4** Blow-Up View of a CD Cross Section*

Unlike the computer hard discs where the tracks are concentric, the CD has a single continuous track spiraling outwards. The spiral track contains billions of very small holes or non-reflective points that are called *pits*. The depth of such pits from the polycarbonate substrate is of the order of 0.12 µm. The areas between the pits are called *lands*.

As the disc spins in the CD-player, a semiconductor laser of approximately 780 nm (i.e., near infrared) wavelength throws light at the reflective surface and by using an optical system that allows the light to travel in both directions to and from the disc surface, it is possible to focus reflected beam onto a detector, a photodiode and pick up the signal. When the beam is reflected from the lands, with no signal when the beam falls on to a pit. The transition from pit to land and from line to pit corresponds to the coding of a 1 in the data stream. A 0 is called as no transition. Only light from laser source can fulfil the requirements of being perfectly monochromatic (one single frequency) and coherent (no break in the wave train) so as to palmate focusing to such a fine spot (see Figure 5.5).

NOTES

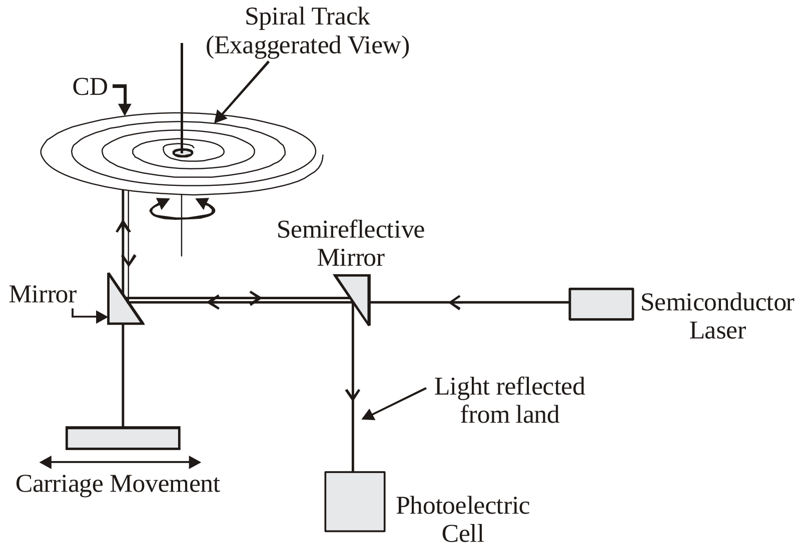


Fig. 5.5 *A Compact Disc*

Compact Disc, Read-Only Memory (CD-ROM)

CD-ROM is another version of the CD that is designed to store both uncompressed hi-fidelity digital stereo sound (like its predecessor) as well as digital computer data in the form of binary files, text, graphics and video. The basic construction of the CD-ROM is the same as for audio CDs (CD-DA), i.e., a standard optical grade polycarbonate substrate 120 mm in diameter and 1.2 mm in thickness with one or more thin reflective metal (aluminium) layers with a lacquer/resin coating.

The original data format was defined and standardized jointly by Philips and Sony in the Yellow Book in 1983. It was a bit too general. Other standards were subsequently introduced and the yellow book was extended. It defines the directory and file structures, including ISO 9660, Hierarchal File System (HFS) for Macintosh computers), and Hybrid HFS-ISO. Today, CD-ROMs are standardized and work in any standard CD-ROM drive to read digital multimedia data (text, graphics, etc.), as well as audio compact discs for music.

Though the disc media and the drives of both the CD and the CD-ROM are basically the same, inside data is stored differently. Unlike the CD-DA, two new sectors were defined:

Mode 1 for storing computer data and Mode 2 for compressed audio or video/ graphic data. Let us inspect them briefly:

CD-ROM Mode 1

For CD-ROMs, which only carry data and applications, the CD-ROM Mode 1 is employed. Many data files are stored on this type of CD, and to access those files an exact address is required to retrieve each file separately. Data is spread out similar to as on audio disks, i.e., sector-wise. Each of which holds 2,352 bytes of data, with an extra number of bytes used for error detection and correction, as well as control structures. The sectors are further broken down for Mode 1 CD-ROM data storage as follows (see Figure 5.6).

- 12 bytes used for synchronization or detection of the beginning of the block.
- 4 bytes for the header, which carries a unique specification of the block.
- 2,048 bytes for the user data.
- 4 bytes for error detection.
- 8 unused or blank bytes.
- 276 bytes for error correction.

NOTES

Sync	Header	User Data	EDC	Blanks	ECC
12	4	2048	4	8	276

2,352 bytes

Fig. 5.6 *Data Block – CD-ROM in Mode-1*

A CD-ROM contains $\frac{6265728000}{8 \times 2352}$ = 333,000 blocks to be played in 74 minutes.

Hence, we can calculate the capacity of a CD-ROM with all blocks in *Mode-1* as:

Capacity CD-ROM Mode-1 =

$$= 333{,}000 \text{ blocks} \times 2048 \text{ bytes/block}$$
$$= 681{,}984{,}000 \text{ bytes}$$
$$= \frac{681984000}{1024 \times 1024}$$
$$\approx 660 \text{ M bytes}$$

Similarly, you can calculate the data rate in Mode-1 as:

$$= 2{,}048 \text{ bytes/Block} \times 75 \text{ Blocks/sec}$$
$$= 150 \text{ Kbytes/sec}$$

This data rate of 150 KB/Sec is called *1X* and was the transfer rate supported in the early CD-ROMs.

CD-ROM Mode 2

The **CD-ROM Mode 2** is used for compressed audio/video information. Here, you might sacrifice occasional loss of a bit of a data as the resultant audio or video output will be barely noticeable. So instead of the elaborate error detection and correction adopted in Mode-1, here the entire 2,336 bytes of data behind the sync and header bytes are used for user data. Also the data is read at 75 blocks/sec rate (see Figure 5.7).

CD-ROM mode 2: Used for compressed audio/video information

Sync	Header	User Data
12	4	2,336

2,352 bytes

Fig. 5.7 *Data Block – CD-ROM in Mode-2*

NOTES

So, the capacity of CD-ROM in Mode-2

= 333,000 blocks × 2336 bytes / block ≈ 741 MB

Similarly, the data rate for CD-ROM in Mode-2

= 2336 bytes/block × 75 blocks/sec ≈ 175.2 Kbytes/s

Both the preceding values are substantially higher than that of Mode-1.

Constant Linear Velocity (CLV)

In the early days of CD, when it was used mainly to reproduce audio, it was imperative to keep the data rate constant (otherwise the music or the voice would quiver). So the manufacturers ensured that the track read by the laser beam moves at constant linear velocity. In other words, the motor driving the CD-spindle rotated at faster speed when accessing data near the centre than when it was near the outer periphery. This technology of rotating the CD at different speed to keep the linear track speed as constant is called ***Constant Linear Velocity (CLV)*.**

Constant Linear Velocity (CLV): Technology of rotating the CD at different speed to keep the linear track speed as constant is called

Constant Angular Velocity (CAV)

The CLV technology was quite satisfactory for the CD-DA or playing audio at 150 Kbytes/sec. However, as the CD-ROM technology evolved to store computer data and the speed increased from 1x to 2X, 4X, 6X, …. 16X, it was found that rotating the CD at the desired speed maintaining CLV to deliver the data was practically impossible and uneconomic. Moreover, audio tracks are generally accessed sequentially, but computer data has to be accessed randomly so the head has to move very quickly from the outer to the inner portion.

To solve the problem, the industry adopted constant speed drives and since the angular velocity is kept constant termed the technology as CAV. It should be noted that the data rate for CAV drives varies depending on the position of the track being read. So the CAV drives are labelled as; variable speed' (e.g., 48X Max). Most CAV CD-ROMs can also read data in CLV mode so that we can still listen the music in a CD-DA disc using a CAV drive.

In between the CLV and the CAV drives, some drives increase the transfer rate till the maximum CAV is reached, then data is accessed from the outward portion of the track by CLV. This is called Partial CAV or P-CAV technology.

Also, some drives uses a technique called Zoned CLV where instead of uniformly decreasing the speed as the head goes from inner to outer, the track is divided into 3–4 zones and for each zone a uniform CLV is maintained. This is termed Z-CLV.

Drive Speed – the X-Nomenclature

The reading speed of CD-ROM/ DVD drives is usually compared with the speed of the original CD player at 150 KB/sec. This speed is adopted as the reference point (1X) and the later generation optical drives are described as multiples of this value. (see Table 5.1)

Table 5.1 Reading Speed of CD-ROM/DVD Drives

CD		DVD			Blu-Ray		
Speed	MB/s	Speed	MB/s	CD equivalent	Speed	MB/s	DVD equivalent
1X	0.15	1X	1.32	9X	1X	4.5	3X
4X	0.6	4X	5.28	36X	2X	9.0	7X
24X	3.6	6X	7.93	54X	4X	18.0	14X
48X	7.2	12X	15.85	--			
52X	7.8	16X	21.13	--			

Source: http://www.osta.org/technology/dvdqa/dvdqa4.htm

NOTES

CD-Interactive (CD-I)

The CD-Interactive or CD-I format was introduced jointly by Philips and Sony in 1986 with a view to develop both a format and a special new type of hardware to access the various multimedia elements interactively including text, graphics, audio, video and computer programs. CD-I represents an entire system. It contains a CD-ROM based format for interleaving of different media as well as definition of compression for different media. CD-I also contained software for real-time processing of media. The CD-I hardware is called the decoder. Its size is comparable to the size of a VCR.

CD-ROM Extended Architecture (CD-ROM/XA)

In the original Yellow Book standard, there was no provision for audio or video data. However, it defined how to store computer data. CD-ROM Extended Architecture (CD-ROM/XA) is an extension of the Yellow Book that introduced two new track types allowing a CD to store computer data (text, binary files, etc.) with compressed audio and/or video data. Here, subheader fields are introduced in both Mode-1 (computer data) and Mode-2 (audio/video data) so that the computer can separate the two types of data on the fly:

Form 1: Similar to CD-ROM Mode 1. Here the unused 8 bytes are used for the sub-header.

Form 2: Similar to CD-ROM Mode 2. Here the unused 8 bytes are used for sub-header and 4 bytes are assigned for error detection, permitting only 2324 bytes for user data.

The CD-ROM/XA formats are not very common except in Kodak PhotoCD, VideoCD and the Sony PlayStation CDs.

CD-Recordable (CD-R)

The CD Recordable (CD-R) and CD Rewritable (CD-RW) drives are called CD burners or CD writers. They are different from the CD-DA or CD-ROM drives in the sense that they have a more powerful laser that besides reading from the discs can also record data to special types CD media (CD-R or CD-RW discs). These CD writers have become extremely popular due to the convenience in duplicating commercial audio and data CDs, or archival of large volume of data. The flexibility, reliability and low cost of both the CD writers, as well as the CD-R and CD-RW discs have made them the most popular PC peripherals – until recently – when the DVD technology started to replace the CDs.

NOTES

Data is recorded permanently in the CD-R discs. The CD-R technology is based on the **Orange Book (Part – II)** standard that was published by Philips in the late 1990.

The CD-R discs have a different structure although the basic geometrical features (120 mm diameter and the drive spindle, etc.) are the same as the standard CD-DA and CD-ROM discs. Like the pressed CDs, the CD-R has the silk-screened label on a protective lacquer layer. Thereafter, comes the reflective layer of aluminium. The CD-R disc, however, has an extra layer between the reflective aluminium layer and the clear optical grade polycarbonate layer. This layer is an organic dye or pigment that is sensitive to light and heat of the CD writer laser of near infrared 780 nm wavelength. Though the various photosensitive dyes have their own colour, they are effectively transparent for the laser beam used. In case of a CD-R, instead of a spiral track having pits molded into the plastic, the spiral track has a groove that has a wobble in it in the form of a sine wave (at the frequency of 22.05 kHz, i.e., half that of the 44.1 kHz sampling rate for audio CDs). The wobble is used to guide the CD-R recording laser beam to write at the correct speed and to follow the groove precisely. When the CD-R is written to, the power of the writing laser is modulated to almost 10 times that of the read power used to read from a disc. The laser operating at write power heats up the disc causing a chemical reaction in the dye that makes it opaque at the locality where the laser hits the track. Thus, the pits are created permanently on the track. When the CD-R disc is read, the light from the reading laser is absorbed and scattered, so that the reading laser recognizes that area as a pit. On the other hand, the unburnt portions of the track allow the laser to be reflected from the shiny reflective aluminium layer and they are recognized as lands.

As already told, the photosensitive dye is altered chemically by the heat of the laser beam and the change is irreversible and permanent. However, some drives allow to record data in multiple sittings provided the disc is not full. This is termed *multi session recording*.

CD-Rewritable (CD-RW)

CD-Rewritable (CD-RW) discs allow data to be erased and rewritten. So a CD-RW can be used repeatedly. CD-RW discs are slightly costly than the CD-R discs.

The CD-RW technology is based on the **Orange Book (Part – III)** standard that was published by Philips.

The CD-RW discs are constructed similarly to CD-R discs. However, instead of using organic photosensitive dye for the recording layer, CD-RW discs use a recording layer comprising of a crystalline compound of silver (Ag), indium (In), antimony (Sb), and tellurium (Te). The alloy when heated to a high temperature and then cooled assumes an amorphous (disordered) form. Otherwise it has a shiny crystalline structure. By heating the alloy to a lower temperature and cooling down, the crystalline structure may be brought back. When the material is crystalline, it reflects more light; so in the crystalline state it is like a 'land' and in the non-crystalline state, a 'pit'. By increasing the power of the laser, it is possible to create the 'pits'. To erase the data the 'pits' are heated by a lower power laser and then allowed to cool. The pit area becomes shiny (Land) once again.

As CD-RW discs are less reflective than CD-R discs and much less reflective than standard CD-Das and CD-ROMs that are manufactured by stamping, many CD-ROM drives and consumer CD players made in the late 1990s or earlier cannot read them.

NOTES

Magneto Optical Discs

Magneto Optical (MO) discs are a special type of discs containing a layer of ferromagnetic material sealed within optical grade plastic. The popular size of the disc is about 3.5″ of capacity 128 MB. Larger cartridges of size 5.25″ with capacity 650 MB to 1.3 GB are also available.

The MO-drives use both optical and magnetic energies to encode data. A laser beam is focused on the surface of the disc. The energy of the beam heats up a tiny spot in the alloy above a critical temperature. The heat loosens the metallic crystals in the alloy enough that they can be moved by the special write head's strong magnetic field. The write head aligns the crystal in one direction to represent a 0 and a different direction to represent a 1. The sensor containing a photodiode senses the orientation patterns of the magnetic signals on the disc, as it changes the polarization of laser light reflected back from the disc. It is a re-writable (RW) technology as the 'pits' created by heating and magnetizing can be reversed by the same process. To read data from the MO Disc, a weaker laser is focussed on the track and the alignment of the alloy crystals representing 1s and 0s reflect light in different ways to the sensor.

LaserDisc (LD)

The LaserDisc was a popular optical disc storage medium for movies in Japan and the USA. Like a conventional CD, the disc contained a reflective aluminium layer with lands and pits stamped on the disc. Video and audio were stored on the LD-CD as frequency modulated signals. While the video signal was stored as analog, the audio tracks were optionally stored as digital signals using pulse code modulation (PCM). Dolby Digital first became available on laserdisc. The laserdisc has now been defunct with the advent of DVD technology, which offers far more capacity and comparable picture and audio quality incorporating both Dolby Digital and DTS at much affordable price.

DVD

DVD is also an optical disc storage media format. Its main uses are video and data storage. DVDs are of the same dimensions as Compact Discs (CDs), but store more than six times as much data. This was achieved by compressing more data using better laser technology and by adding additional layers in the disc.

DVD: An optical disc storage media format namely used for storing video and data

In the initial period of introduction, DVD was known as 'Digital Video Disc' and thereafter when it was used to store data as well, as 'Digital Versatile Disc'. Subsequently, DVD Forum, the official body for DVD standards, clarified that DVD is not an acronym, i.e., the three letters D,V and D does not stand for anything.

Every DVD disc should use the same physical file structure, promoted by the Optical Storage Technology Association (OSTA), and called Universal Disc Format (UDF). Any DVD drive can read any file from any DVD disc. This has at least removed the incompatibility of formats to a large extent that prevailed with CDs.

NOTES

Variations of DVD, however, exist in terms of capacity and the way data is stored on the discs. The implementations range from single sided DVD discs (1.46 GB) to multi-layered DVD-18 disc (double sided, 4 layers) having more than 17 GB capacity.

Table 5.2 shows the comparative status of various types of DVD.

Table 5.2 *Comparative Status of Various DVD Types*

Type	Spec.	Diameter	Total number of Layers	Sides	Capacity in GB (billion of Bytes)	Actual GB (2^{30})
DVD-1	SS/SL	80 mm	1	1	1.45	1.36
DVD-2	SS/DL	80 mm	2	1	2.65	2.47
DVD-3	DS/SL	80 mm	2	2	2.9	2.72
DVD-4	DS/DL	80 mm	4	2	5.3	4.95
DVD-5	SS/SL	120 mm	1	1	4.7	4.38
DVD-9	SS/DL	120 mm	2	1	8.5	7.95
DVD-10	DS/SL	120 mm	2	2	9.4	8.75
DVD-14	DS/SL+DL	120 mm	3	2	13.24	12.33
DVD-18	DS/DL	120 mm	4	2	17	15.90

A single layer–single sided disc has only one substrate with a reflective surface and a data layer with a blank substrate. This is the DVD-5 or DVD-1 format. The total thickness is 1.2 mm. On the other hand, a single layer double sided disc is formed by bonding together two single sided substrates back to back each 0.6 mm thick. This is the DVD-10 or DVD-3 format.

The DVD standard also permits two layers in a substrate, one below the other, resulting in a dual layer disc—the DVD-9 or DVD-2 format.
Similarly, each substrate having two layers when pasted together back-to-back results in a double-sided double layer disc—the DVD-18 or DVD-4 formats.

The schematic representations are shown in Figure 5.8.

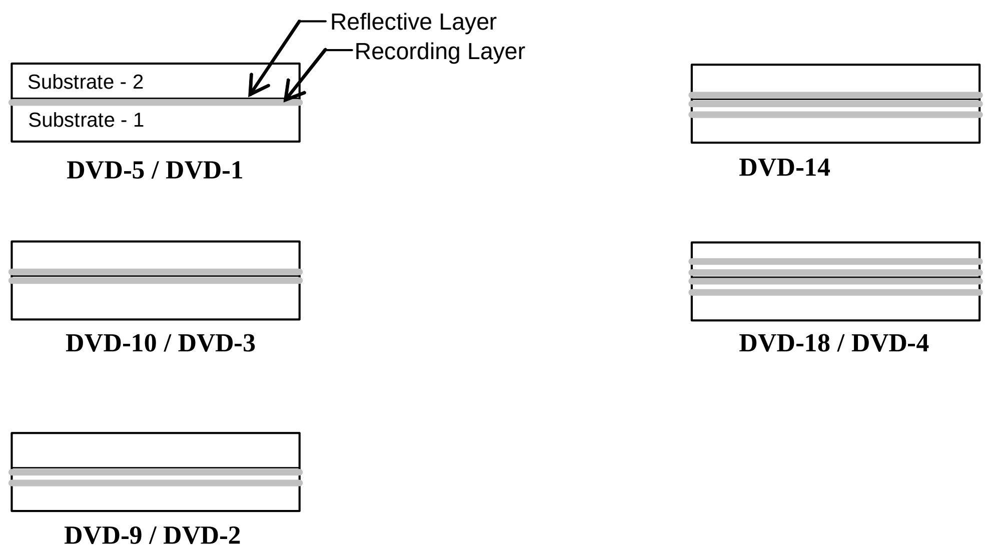


Fig. 5.8 *Schematic Representations of DVD*

To read the different layers on the same substrate, the reading Laser has to change the focus. To read the data from the other side, the DVD has to be taken out of the drive and manually turned over.

DVD-ROM Drive Speed

NOTES

Just like the CD drives, the speed of the DVD drives is specified by the X-factor, i.e., 1X, 4X, 6X, etc. However, a 1X CD drive transfers data at 150 KB/sec, while a 1X DVD drive transfers data at 1321 KB/sec, i.e., about nine times the 1X CD rate. The comparative chart has been given in Table 5.1 (Under Drive Speed—the X-nomenclature). For DVD-video, the file is accessed sequentially, and it always plays at 1X, but for DVD-ROM the transfer rate or throughput is important while accessing multimedia and other digital data.

Due to the difficulties in maintaining CLV at the extremely high transfer rate, the current DVD-ROM drives use CAV. You will notice the DVD drive mentions the speed rating as the 'Max'. The disc spins at a constant speed for a particular zone, which gradually increases as the read laser goes from the inner to the outer portion of the spiral track. The maximum speed is achieved when the laser reads from the outermost track. DVD-ROM is made by stamping the lands and pits on the polycarbonate substrate in the manufacturing plant.

DVD Formats

The following are the various formats of DVD:

DVD-Video

In marketing parlance, there exist two types of DVD—DVD-Video and DVD-ROM. DVD-Video is the standard for delivery of video content on DVD media. Today, DVD-Video has almost totally replaced the VHS cassettes for distribution of video in the worldwide market. The DVD-Video supports many formats and resolutions.

The consumer DVD-Video discs normally use 4:3 aspect ratio (or anamorphic 16:9 ratio) MPEG-2 video, at the resolution of 720 × 576 (PAL) or 720 × 480 (NTSC) at 29.97, 25 or 23.976 frames per second (FPS). Audio is stored using the Dolby Digital (AC-3) or Digital Theater System (DTS) format in either 16-bits/48 kHz or 24-bits/96 kHz format with mono to 7.1-channel *Surround Sound* mode, and/or MPEG-1 Layer 2. As already discussed, the specifications for video and audio vary by regions across the world depending on the television format used, however, often the DVD players support all the formats. DVD Video also supports features, such as subtitling, menus, multiple audio tracks, camera angles, etc.

DVD-Audio

DVD-Audio offers higher fidelity than CD-quality, with sampling rates of 44.1, 48, 88.2, 96, 176.4, and 192 kHz and a variety of sample sizes. Audio compression options include MPEG and Dolby Digital (AC3). DVD-Audio discs have the option to apply a copy protection mechanism, termed *Content Protection for Prerecorded Media (CPPM)*, which was developed jointly by IBM, Intel, Matsushita and Toshiba. It is, however, debatable whether the acoustic fidelity claimed to be achieved by sampling above 44.1 kHz is at all distinguishable to a human listener.

DVD-Audio: Offers higher fidelity than CD-quality, with sampling rates of 44.1, 48, 88.2, 96, 176.4, and 192 kHz

NOTES

DVD-R

The DVD-Recordable (DVD-R) was initially designed as a recordable (write once) disc having more capacity than the CD-R format so that it could compete with the popular VHS videocassettes. The working principle is similar to that of a CD-R. Information is recorded on a groove in a layer of photosensitive organic dye. Once the laser alters the dye, the information is permanently etched on the track and cannot be changed. The Single Surface/Single Layer (SS/SL) disc is manufactured by bonding two polycarbonate substrates, each 0.6 mm thick (and 120 mm in diameter), one having the reflective aluminium coating over the photosensitive organic dye layer, and the other simply the polycarbonate disc. The capacity of such a DVD is 4.7 GB.

Though the CD and DVD are of same size and thickness yet there are some basic differences. For a DVD, the laser beam has to penetrate only 0.6 mm of the polycarbonate substrate, while for a CD it has to penetrate about 1.2 mm. Also for a CD a 780 nm red laser is used as the reading laser; while for a DVD-R the reading laser is a shorter wavelength 650 nm red laser resulting in smaller and more compact pits on the track. Further, the pitch of the tracks is also reduced (than that of the CD) to accommodate more tracks in the data area. In a DVD-R, the correct location of the laser beam on the grooved track is ensured by special markings called land pre-pits.

The DVD-R (pronounced as 'DVD dash R') is approved by DVD Forum (www.dvdforum.org)—the international organization set up to standardize DVD formats. Initially, the DVD format and logo license fees fixed by the DVD Forum was very high. So a rival group of manufacturers came out with the DVD+R and DVD+W with almost identical specifications. The Hybrid drives can handle both DVD+R and DVD-R discs, and they are labelled as DVDRW drives.

DVD+R

The DVD+Recordable (DVD+R, pronounced: DVD plus R) is also a write once optical disc of 4.7 GB capacity. The format was floated by the DVD+RW alliance (www.dvdrw.org) in 2002.

The DVD+R DL or Dual Layer format (also termed DVD+R9) was introduced by the DVD+RW Alliance in 2003. This format almost doubled the storage capacity (from 4.7 GB to 8.4 GB) by using two layers of photosensitive dyes instead of one. Specialized DVD+RW DL drives are however required to write and read from these DL discs.

DVD-RW

The DVD-Rewritable (DVD-RW) discs allow data to be erased and rewritten. So a DVD-RW can be used repeatedly (about 1000 times). The DVD-RW format was introduced by Pioneer as per the approved standards of DVD Forum (the DVD-RW book). However, the DVD-RW has an inherent limitation. The DVD-RW was introduced as an alternative to the rewritable DVD-RAM since the video recordings on DVD-RAM could not be played on regular DVD players. So the industry wanted an erasable DVD medium to be reused just as a videotape. The DVD-RW perfectly fits in that role. However, when used for data, it has a problem. The DVD-RW format of storage of data is sequential in nature just as video recordings are essentially sequential and a new video is normally appended at the end of previous recordings. This sequential

design, however, stops data from being erased from the DVD-RW to permit more room in their place. Deleting files from a DVD-RW do not increase the capacity of a DVD-RW disc. This problem has been addressed and solved by the DVD+RW alliance in their DVD+RW discs.

NOTES

The DVD-RW discs are constructed similarly to the DVD-R discs. However, like the CD-RW discs, the DVD-RW discs use a recording layer made of a special alloy. The alloy, when heated to a high temperature and then cooled assumes an amorphous (disordered) form. Otherwise, it has a shiny crystalline structure. By heating the alloy to a lower temperature and cooling down, the crystalline structure may be brought back. When the material is crystalline, it reflects more light; so in the crystalline state it is like a 'land' and in the non-crystalline state, a 'pit'. By increasing the power of the laser, it is possible to create the 'pits'. To erase the data the 'pits' are heated by a lower power laser and then allowed to cool. The pit area becomes shiny (land) once again.

DVD+RW

The DVD+RW is another type of rewritable DVD, which was introduced by the DVD+RW alliance and given the name 'DVD+RW' to differentiate it from the DVD-RW (pronounced: DVD dash RW). It can store both video/audio and data effectively. It can introduce the error management systems when handling computer data and can dispense them with when storing audio/video so that standard DVD players can recognize the discs. The basic geometry of a DVD+RW disc, such as diameter, thickness, etc., is the same as a DVD-RW or a CD-RW disc.

You have already learned that the DVD-RW discs are suited for storing audio/video sequential data but not so for storing data. The 'DVD+RW Alliance' had observed the shortcomings of the DVD-RW format and introduced a new formatting structure giving some significant advantages over the DVD-RW. As a result, the DVD+RW became popular for storing both computer data as well as video/audio. The structure is based on an 817.4 kHz sine wave like wobbly groove moulded into the polycarbonate base. This wobble is about 37 times finer than the wobble in a CD-R, and the groove with the wobbly waveform makes aligning data blocks much convenient and accurate, because it serves as the disc's marker guide for the laser beam to address the correct position on the track. Also adding to or erasing information from a DVD+RW disc could be done with more accuracy and speed. For example, the time taken to fully erase a DVD-RW disc takes more than 80 minutes, whereas for DVD+RW disc it takes less than a minute and that too as a background process. Further, the total time to record a DVD+RW disc is less.

DVD-RAM

This **DVD-RAM** is a disc specification designed for Random Access Memory (RAM) to access data very quickly. DVD Forum introduced the specification in 1996 for both the DVD-RAM drive as well as DVD-RAM media. DVD-RAM is a reliable storage medium for computers and Camcorders. It can be used as a data storage and back-up medium and also to record video.

DVD-RAM: A disc specification designed for Random Access Memory (RAM) to access data very quickly

The DVD-RAM disc has grooves moulded into a polycarbonate substrate, which is bonded to a second polycarbonate disc. It has predefined pits to identify

NOTES

sectors for address information so that a drive can very quickly locate files. These moulded addresses or pits are visible on the DVD-RAM discs as a series of small lines. Unlike the recordable or rewritable DVDs, recording in a DVD-RAM is done in both the Land between the grooves as well as in the grooves. The DVD-RAM is enclosed in a protective cartridge. DVD-RAM was designed to handle computer data and be more rugged. It has superior defect management, uses zoned CLV (i.e., Partial Constant Angular Velocity or PCAV) for faster data access, and offers greater media protection by means of a cartridge. However, the drawback is that the earlier DVD-ROM drives and almost no DVD player can read from the DVD-RAM discs. DVD-RAM may be the best choice if DVD writer is dedicatedly used only to back up or archive computer data. It is very reliable, lasts long (about 30 years) and A DVD-RAM disc can be rewritten at least 1,00,000 times. The write speed of DVD-RAM drives is much less than that of DVD-RW and DVD+RW drives and the price of DVD-RAM discs is more than DVD-RW and DVD+RW discs.

Hitachi, Toshiba and Panasonic (Matsushita) support the DVD-RAM standard. The two variants of DVD-RAM are:

First-generation (DVD-RAM Book 1.0)**:** It records 2.58 GB per side on rewritable media. These discs are not readable by older DVD players and drives.

Second-generation (DVD-RAM Book 2.1)**:** It reads and writes both original 2.58/5.2 GB DVD-RAM discs (first generation) and 4.7/9.4 GB DVD-RAM discs. Non-cartridge 4.7 GB and 9.4 GB DVD-RAM discs are now widely available. However, older DVD-RAM drives often do not support them.

5.3 MEDIA TYPES

Disk Drive: a peripheral device used to store and collect information

A **disk drive** is a peripheral device used to store and collect information. It can be removable or fixed, high capacity or low capacity, fast or slow speed, and magnetic or optical.

Structurally, a drive is the object inside which a disk is either permanently or temporarily stored. While a disk contains the media on which the data is stored, a drive contains the machinery and circuitry required for implementing read/write operations on the disk.

The disk looks literally like a flat circular plate. The computer writes information to the disk, where it is stored in the same form as it is stored on a cassette tape. Disks as such are just magnetically coated rolls or circular disks which are divided into sectors and tracks. The data is accordingly stored and numbered with respect to the track and sector number on the disk; only the structure of the medium is different. Examples of removable disk drives are DVD, CD-ROM, floppy disk drive, etc. A hard disk is an example of a non-removable disk drive.

The method of accessing data could be sequential (magnetic tape drives) or random (HDD, DVD), where the read/write head can directly go to any location on the disk.

Drives can be classified into two major groups – magnetic and optical. The following section gives a brief overview of various types of magnetic and optical drive.

Magnetic Drives: These are magnetized storage media on which digital or analog information is recorded as electromagnetic signals over tracks and sectors predesigned on the media. They are a non-volatile source for storing data because they can store information for a long time and do not require electricity or any other element to retain the information stored in them.

Hard Disk Drive: This is a crucial hardware component of a personal computer, without which modern-day computers cannot function. Although the RAM is a place of primary storage, it is ephemeral (i.e., its life is dependant on the power source – RAM is active only as long as the computer is turned on). A hard disk drive, though technically a secondary form of storage, is the primary form of permanent storage (since the data stored on the hard disk is not dependant on the computer being switched on or off). The features of HDDs that have made them an irreplaceable component in our computers are their high capacity for storing data and the high speeds at which they can access it at relatively lower cost. They come in various interfaces and specifications such as IDE, EIDE, SCSI, SATA and SATA II.

Floppy Disks: These are portable media consisting of a magnetically coated disk kept inside a protective covering. They are low capacity and cheap to manufacture, but highly prone to dust and scratching. Due to these limitations, they are no longer considered a standard component of a personal computer system. Their size varies from 360K to 2.88MB.

ZIP Drives: These are similar to disk drives but with thicker magnetic disks and a larger number of heads in the drive to read/write. The Zip drive was introduced mainly to overcome the limitations of the floppy drive and replace it with a higher capacity and faster medium. They are better than floppy disks but still slow in performance and with a high cost-to-storage ratio. The disk size ranges from 100MB to 750MB. Zip drives were popular for several years until the introduction of CD-ROMs and CD-Writers, which have now come to be widely accepted due to their cost, convenience and speed.

Tape Drives and Tape Drums: Tape drives represented the sequential access method of storage and retrieval. Sequential access, as opposed to the now prevalent random access, means that data can be stored and retrieved only in a sequential manner – as defined by the order in which the data was stored on the tape drive. For instance, in case of cassette tape drives (used in tape recorders and players), which are an example of sequential access based storage, if you want to play song number 4, you can do so only after you have either played or fast forwarded song numbers 1 to 3 (there is no way to *directly* go to song number 4 without going through songs 1 to 3). Tape drives were widely used in the 1980s and 1990s as backup devices, but due to their slow speed and sequential read/write access, they have now become virtually obsolete.

Optical Drives

An **optical drive** is a type of storage medium that stores the content in digital form which is written and read by a low intensity laser. The laser reads data from the reflective surface of an optical disc by measuring surface changes in height and depth. All types of optical media are divided into tracks and sectors that contain a series of tiny indentations in which data is stored.

Optical drives: A type of storage medium that stores the content in digital form which is written and read by a low intensity laser

NOTES

CD-ROM: This is an optical medium of data storage. The current maximum capacity of a CD-ROM is 900MB with a maximum read/write access speed of 52X, (which means 10,350 RPM - rotations per minute) and transfer rate of 7.62 MBPS (mega-bytes per second). The data is written with the help of a red infrared laser beam from an optical lens and the same laser of lower intensity is used to read data from the CD-ROM.

HD-DVD: A high density, mostly single sided, double layered optical disc which can hold up to 15GB on a single layer and 30GB on a dual layer disc. The read/write speed on an HD-DVD varies between 36 MBPS and 72 MBPS. These were primarily designed for the storage of high-definition videos and large volumes of data. The basic look and feel of an HD-DVD drive and disk is the same as that of a CD-ROM and DVD except that it uses a laser of different wavelength and the microscopic structure of storage on a disk is different.

Blu-Ray: Another high-density optical storage media format is gaining popularity these days. It is mainly used for high-definition video and storing data. The storage capacity of a dual layer Blu-ray Disc is 50 GB, almost equal to storing data in six double-dual layer DVD or more than 10 single-layer DVD.

5.4 MEDIA ORGANIZATION

Recent advances in compression, storage and communication technologies have resulted in the creation of applications that involve storing and retrieving multiple data types, such as text, audio, video, imagery, etc., collectively referred to as *multimedia*. These applications require the development of file systems that can efficiently manage the storage and retrieval of multiple data types referred to as *integrated multimedia file systems*. Earlier, the handling and organization of these specific file types was a cumbersome task when the original and conventional file systems was arranged to accommodate them. With the increasing trends of graphical and media files embedded e-mail windows the handling of multimedia files has become inevitable.

The Multimedia File System (MMFS) was specifically designed to provide a high performance network interface for storing and retrieving multimedia data. The performance optimizations in the MMFS implementation or operating system will not necessitate modifications in any application code. Typically, the manipulation of digital audio data within MMFS is facilitated through various audio-specific programming layers. The MMFS maintains association between related files and also helps in storing multimedia data types, such as MIDI files, still images and video animation frames with a universal standard. MMFS is intended to support continuous media intensive applications, such as personal video recorders, video JukeBoxes and Video-on-Demand (VoD). It completely replaces the VCRs because it works as same as video player works, for example it provides *STOP, PAUSE, FORWARD* and *REWIND* services. For this, it frequently requires Set-Top-Box (STB). The services collectively represent the virtual multimedia and video content shop. A basic concept of directories and files you must know at this stage so that you can get aware of the multimedia file system and information representation. The hierarchical file naming systems contain the following interfaces with reference to multimedia file system:

A Tree of Directories and Files: Contains file and directory names that are named with path names.

Objects: Deals files and directories along with other kinds of objects, e.g., devices.

Path Names: Contains a component name for each directory in the path.

A file is a passive container of bytes on the disk, whereas the ***Open*** file is active source or sink of bytes in a running program. These files are connected to a device or a process involved in OS. In this design technique, devices are named as files that are opened and used in byte streams. The byte streams represent the sources and sinks of bytes.

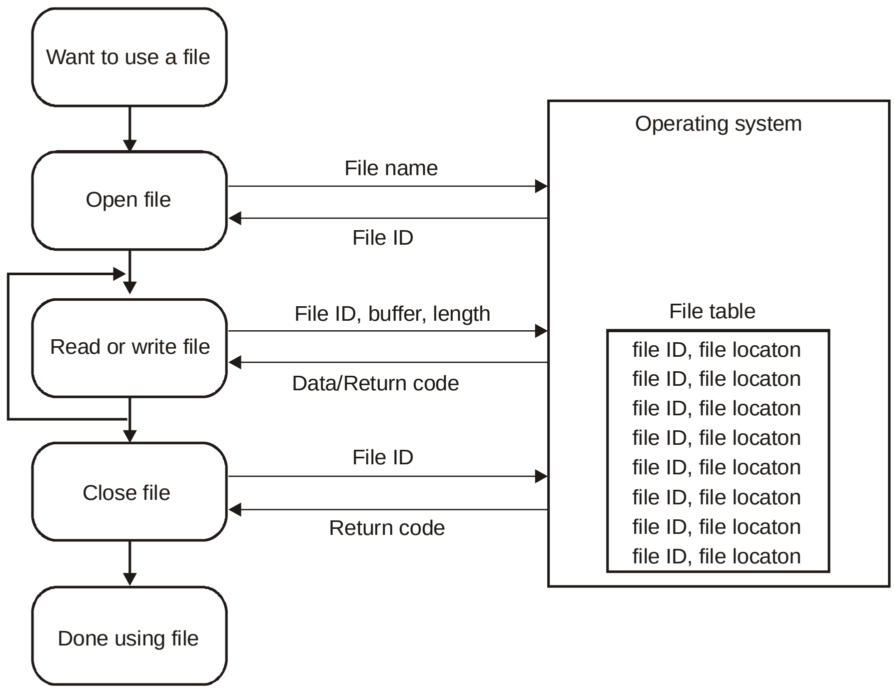


Fig. 5.9 *Files and Open Files*

In Figure 5.9, file ID and file locations are required in the file table that is situated in the operating system and is supported by MMFS. MMFS offers a set of functionalities for multimedia support that is synchronized multi-stream retrieval and used for editing support. This file system also supports caching and prefetching optimizations for real-time disk scheduling. The multimedia file is associated with unique mnode that contains the metadata of the MM file. The multimedia-specific metadata of each strand is used for recording rate, logical block size and the size of the application data unit.



Fig. 5.10 *MMFS Design Prefetching*

Figure 5.10 shows the playback of a video in fast-forward and UFS prefetching, which issues the read-ahead for unnecessary blocks. Frame *K* is reserved for *i*, *i+1*,

NOTES

i+2, whereas Frame *K+3* retains *i+9, i+10, i+11* blocks. Multimedia unique demands in file system and MMFS are sued to extend the Unified File System (UFS). UFS is used to write and truncate the system calls for small size multimedia files. The frame supports in MMFS designing by single medium editing and multiple-media playback environments. A fully functional file system is based on the Virtual File System (VFS). The playback of a video is kept in fast-forward in which MMFS perform intelligent prefetching. The applications communicate MMFS setting the fields in mminfo retrieval rate, direction, whether frames skip the degree of prefetching is maintained at a high level. It does not work for compressed data streams. The challenge is multimedia systems faced in need to replay media types continuously, i.e., the data that should be played must to arrive in real time or at least certain strict deadline. Continuous media data differs from discrete data but not only in its real time characteristics. Also, a challenge for these systems is the synchronization of pictures and the corresponding sound. Therefore, these can be considered as two different data streams. It is important to synchronize these before displaying them on the monitor. Another difference typical to discrete data is the file size. The video and audio need much more storage space than text data and the multimedia file system has to organize this data on disk in an efficient way that utilizes the limited storage. MMFS is designed to support the recording and playback of data streams at constant and variable rates. The following features are available for multimedia file system:

- MMFS provides the Personal Video Recorder (PVR) functionality allowing several data streams to be recorded simultaneously while also replaying a stream, which may be one of the streams being recorded. Provide the ability to fast-forward and rewind data streams.
- MMFS makes efficient use of disk storage, access times and bandwidth.
- MMFS enables the automatic recovery of disk data structures on restart after a power failure or other interruption. Automatic formatting of a new disk or one that is irretrievably corrupt.

MMFS is accessed through the FILEIO package which presents a standard POSIX compatible IO interface through which applications use standard open(), read(), write() and close() calls. Streaming support is provided through a small library, mmfslib, which presents a more application-friendly interface. MMFS supplies most of the standard file I/O functionality. However, since it is optimized for supporting streamed data, it has a number of restrictions that mean that it does not always behave like a general-purpose file system by the following means.

- Files may not be resized after creation and are essentially write-once/read-many. Between the initial open() and close() that creates a file it will be extended as requires. On subsequent opens, even those that specify O_WRITE, data may only be written to the existing file extent.
- If an attempt is made to create a file that already exists, the open() will fail. Instead the file must be deleted first and may then be created anew.

The formatting options control the formatting of an MMFS disk. They are only used when a file system is formatted. Under normal circumstances the file system will fetch these values from the disk volume label. Table 5.3 shows the formatting options of MMFS:

Table 5.3 Formatting Options of MMFS

Options Available for Formatting the MMFS	Function
CYGNUM_FS_MMFS_BLOCK_SIZE	This option defines the size of file system blocks. The value is defined in KiB (kibibyte) and must be a power of 2 and default value is 256.
CYGNUM_FS_MMFS_ROOTDIR_SIZE	This option defines the size of the root directory in blocks. Since all files are contained in this directory, its size gives a hard limit to the number of files that the file system may contain and the default value is 1.
CYGNUM_FS_MMFS_BAT_SIZE	This option defines the size of the block allocation tables (BAT) used to store the addresses of file data blocks. This gives a hard upper limit on the size of a file and the default value is 2.
CYGNUM_FS_MMFS_BAT_COUNT	This option defines the number of BATs allocated in the file system. The default is to define 200 BATs and the default value is 200.

NOTES

The intelligent virtual reality software system includes various types of including multimedia software, VRML etc., whereas the multimedia software mainly uses Adobe Authorware, Adobe Director, Winamp, etc. VR applications are generic software systems that provides user toolkits and APIs, such as OpenGL, ray-tracing systems etc. It is possible to embed videos into html documents, which are known as Web pages in two ways. One method is to use the `<embed/>` tag to display your media file. The embed tag does not require a closing tag. In fact, it works much like the image tag. A `src` attribute is defined by the correct URL, either taken as a local or global in order for correctly displaying the video file. You may start and stop your movie files by either pressing the buttons or double-clicking your mouse (continue/play). It is also possible to place the URL of your media files into the `href` attribute of an anchor tag. Table 5.4 shows the various video media files:

***Table 5.4** Video Media Files*

File Name Extension	***Function***
.swf files	These file types are created by Macromedia's Flash program
.wmv files	These file types are Microsoft's Window's Media Video file types.
.mov files	These file types are Apple's Quick Time Movie format.
.mpeg files	These file types are set the standard for compression movie files created by the Moving Pictures Expert Group.

The flash movies (.swf), AVI's (.avi) and MOV's (.mov) file extension types are supported by the embed tag. Using Adobe Flash CS4 software, a sound file is inserted

NOTES

on Web page that plays after clicking the button. The required steps are accomplished to perform the task:

- Open a new flash page that imports the sound file, let say sound_car.mid, it is imported from the provided library. The overall movie file developed in flash is taken as .swf (shockwave format) extension name. The extension name .mid is sound file created in flash installed in the Internet explorer.
- Select **Window→Common Libraries→Buttons** menu that provides a list of buttons symbols. These buttons are included during the installation of flash CS4.
- Double click on button to open symbol editor which adds frames.
- Click on Scene that links to the main timeline.

The audio files playing music are musical instrument digital interface (midi), .mid, .mp3 and .wav files placed on web page. After knowing the basic concept of multimedia file system, the information is represented by a number of test programs. Table 5.5 shows the number of test programs and their functions:

Table 5.5 *Number of Test Programs and their Functions*

Test Programs	*Functions*
mmfs1	This test program just tests the standard FILEIO interface of the file system. It is a simplified version of the file system functionality tests used by FATFS and RAMFS.
stream1	This test program is a simple test of the streaming support. It writes and reads streams at defined data rates and checks that the rate is maintained and that data integrity is preserved.
pvr1	This test program is a basic emulation of a personal video recorder. Two streams are written and one of them read back, after a delay. This simulates a PVR recording one channel while using a pause-live feature on a second channel.
pvr2	This test program records a large number of short streams on the disk and also checks directory handling.
pvr3	This test program is a variant on pvr1, which records 3 streams while replaying one.
format	This refers to simple test that uses the 'mmfs.format' file system instead of 'mmfs', thus reformatting the disk.
example	This contains versions of the write_stream() and read_stream() example functions described earlier, together with sufficient infrastructure to allow them to be run.

Information Representation in Multimedia Applications

Fourier analysis can be used to show that any time-varying analog signal is made up of a possibly infinite number of single-frequency sinusoidal signals whose amplitude and phase vary continuously with time relative to each other signal bandwidth. The bandwidth of the transmission channel should be equal to or greater than the bandwidth of the signal-band limiting channel. Mathematician Fourier introduced the infinite sum of sine and cosine waves. The harmonic musical sound is produced with audio files. With the help of Fourier analysis, the sound synthesizing is analysed from a series of pure tone generators. It is adjusted by their amplitudes and phases and then by adding them

together. The sustained sound in multimedia files can be reproduced with the limited range of frequencies. Most of the sound energy follows the harmonics of fundamental pitch. This analysis is generalized to interpret that any sound with a sharp attack, sharp pulse and rapid changes can be reproduced in a wave form. It produces in the term of square wave (see Figure 5.11), which contains only odd harmonics with the amplitude.

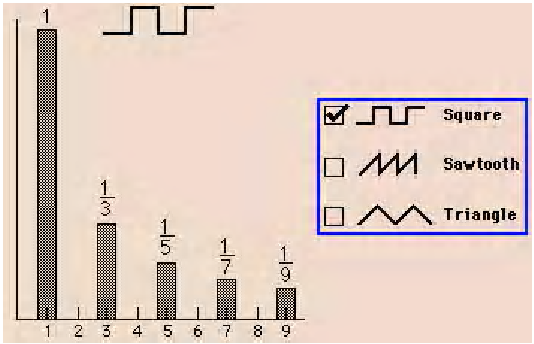


Fig. 5.11 *Square Wave*

The video and broadcast channel makes scanning sequence and it is necessary to use a minimum refresh rate of 50 times per second to avoid flicker. A refresh rate of 25 times per second is sufficient for the transmission of documents. The two fields are then integrated together in the television receiver using a technique known as interlaced scanning. The three main properties of a color source are taken as brightness, hue and saturation. Hue represents the actual color of the source of multimedia file and saturation represents the strength or vividness of the color. Multimedia information representation supports CODEC that performs the conversion using some code words. Coder/ Decoder (CODEC) is a piece of software or a driver that adds a support for certain audio-video format for operating system assembled PC. With CODEC, your system recognizes the format the codec is built for and allows you to play the audio-video file *(=decode)* or in some cases, to change another audio-video file into that format *(=(en)code)*. For example, when you install Windows to your home computer, Windows installs automatically bunch of most commonly used CODEC into the system, so you do not have to download them separately from their vendors.

Table 5.6 *PC Video Digitization*

Digitization Format	**System**	**Spatial Resolution**	**Temporal Resolution**
4:2:2	525-line 625-line	$Y = 640 \times 480, C_b = C_r = 320 \times 240$ $Y = 768 \times 576, C_b = C_r = 384 \times 288$	60Hz 50Hz
Siebel Import File (SIF)	525-line 625-line	$Y = 320 \times 240, C_b = C_r = 160 \times 240$ $Y = 384 \times 288, C_b = C_r = 192 \times 144$	30Hz 25Hz
Common Intermediate Format(CIF)		$Y = 384 \times 288, C_b = C_r = 192 \times 144$	30Hz
Quarter Common Intermediate Format (QCIF)		$Y = 192 \times 144, C_b = C_r = 96 \times 72$	15/7.5Hz

NOTES

For all digitization formats (see Table 5.6), video capture board and software are required. All the PCs use monitors for presenting the progressive and non-interlaced scanning. The clustering is used to produce candidate trajectories in video. Each motion trajectory is described using chain code and is represented by object-motion-video triplets. In audio characterization, audio information is classified into different categories, such as silence, speech, music and other voice. In general, you can distinguish different ways of analysing and searching video:

Table 5.7 displays the way of information representation computer graphics array.

Table 5.7 *Information Representation in Computer Graphics Array*

Standard	Resolution	Number of Colors	Memory Required Per Frame (Bytes) of Multimedia Documents
VGA	640×480×8	256	307.2 KB
XGA	640×480×16 1024×768×8	64K 256	614.4 KB 786.432 KB
SVGA	800×600×16 1024×768×8 1024×768×24	64K 256 16M	960 KB 786.432 KB 2359.296KB

5.5 APPLICATIONS OF MULTIMEDIA

In the twenty-first century, IT provides many services like airlines, hotel management, Web publishing etc. Some of these services are explained in the following sections.

Airlines

The air travel industry is one of the biggest users of information technology. There is hardly any aspect of the airline business in which computer systems have not been deployed for increasing revenues, reducing costs and enhancing customer satisfaction.

It is now almost inconceivable to book a ticket or get a seat confirmed across multiple sales counters (airline offices, travel agents, etc.) spread over numerous cities, without using computerized databases and e-networking. Like most other industries, the use of computerized systems in the air travel industry started with the front office and sales desk with back-office operations playing a oracial role in delivering a quality experience to consumers. What typically started as airlines intranet systems have now blossomed into vast Web-based online systems which can be accessed by anybody from anywhere in the world.

The following are some of the interesting areas where IT has been used successfully:

1. **Online Ticket Reservation Through the Internet:** Today, most leading airlines like United Airlines, Delta, British Airways, etc. sell tickets through their Websites. You can book the ticket through the Internet, pay online by giving your international credit card details and then collect the ticket (on the day of journey), boarding pass from e-ticket kiosks at the airport by simply furnishing your booking reference details.

Check Your Progress

1. What is a CD?
2. How CD recording process takes place?
3. What is special about CD-Rewriteable?
4. What is a disk drive?
5. What do you mean by an optical drive?
6. Name some of the areas in airlines where IT has been helpful?
7. Why 'infotel' is used?
8. What is a cellular phone?
9. What is Web publishing?
10. What is the use of bar coding technique?

NOTES

2. **Flight and Seats Availability:** If you wish to travel from New Delhi to New York and do not know what your flight options are, simply log onto the airline site, specify the cities of travel origin and destination along with preferred journey dates and the database would yield all the possible options. Once you have selected the flights, you could even go a step further (possible in case of a few airlines) and book a specific seat number in that flight along with the choice of meal.
3. **Last Minute Deals and Auctions:** A seat is a perishable commodity. An unsold seat means a revenue opportunity lost forever. Therefore, most airlines have now started a facility on their Website where potential customers can bid for last minute tickets in online auctions. Cases of people buying a ticket worth $1000 for as low as $100 are not uncommon. This is a case of win-win by effective use of IT—the passenger is happy at getting the ticket at a fraction of the ticket's normal cost and the airline is able to recover something from what might otherwise have been an unsold seat.

All these facilities/opportunities would have been impossible without an integrated online computer system.

Telephone Exchanges

The first telephone service invented by Alexander Graham Bell was strictly 'point-to-point', i.e. each user had to be physically wired to every other user. There was no 'telephone exchange'. Needless to say, Bell immediately realized the need for an exchange and made one. In this first exchange each subscriber had to be wired only up to his local exchange. An operator sitting in the exchange, connected him to other subscribers upon request (earlier phones did not have dialing facility) by physically connecting the caller's wire to the recipient subscriber's telephone by using a hand-actuated circuit switch. One does not need to stretch one's imagination to appreciate the fact that operator-controlled exchanges were not only extremely labour-intensive but also highly error-prone.

Now, compare this to the digital, computerized telephone exchanges used today. These are the electronic systems that do the switching operation based upon a 'stored program control.' The rules defined in the software assess which destination the caller is trying to reach, plot the most optimal path, intimate the called party, inform the caller about his call status and then if the called party accepts the call, establish the circuit. The call is monitored during its progress and the circuit is disconnected once the call is terminated. Computerized exchanges improve and enhance call-processing capacity, thereby lowering the cost of operations. They also opened up a wide array of IT-enabled services for subscribers that have made modern telephony an indispensable service.

Bharat Sanchar Nigam Limited (BSNL), one of the main providers of telephone services extensively use a product called Infotel for managing their telephone exchanges. This product provides:

1. **Provision of Facilities:** Activation, deactivation and modification of subscriber facilities, such as ISD, STD, call waiting, call transfer, computer-generated bills, etc.

NOTES

2. **Fault Booking and Restoration Service:** Maintenance of a database of complaint calls either through an IVRS (interactive voice response system) or a customer service cell. The system automatically creates the complaint docket and generates a range of statistical and exception reports.
3. **Line Data Maintenance:** The system provides online data on cable codes, cable pair numbers, cabinet number, pillar numbers, etc. for all subscriber connections to facilitate and expedite line repair and maintenance.
4. **Directory Enquiry:** The computerized subscriber database also allows extensive online or voice-based directory enquiry based upon subscriber name, location, telephone number etc.

Mobile Phones

Statistically, major portions of the population of any developing country still do not possess a telephone. Making a simple call to anybody requires locating the nearest telephone booth, waiting for ones' turn in the queue and then paying for a short chat on (most often) a disturbed line.

In the developing countries, the penetration of landline phones has been low largely due to the hassles of laying cables across long distances. Especially in the case of remote areas, the cost of connecting a few phones to the mainland mass becomes disproportionately high. Maintaining these telephone cables across inhospitable terrain also poses a major challenge to network expansion planners and engineers.

Due to the advances made in the telecommunications industry in the last two decades, mobile phones provide an excellent cost-effective and efficient alternative to the land phones for developing countries like India.

A cellular phone (as mobile phones are also known) is primarily a radio—a very sophisticated variant of a radio telephone. The genius of a cellular system is the division of the city into small cells (hexagons on a big hexagonal grid). Each cell has a base station that consists of a tower and a small building containing the radio equipment. Wireless communication is possible within and across the cells, allowing a user complete mobility and making communication much easier and less time-consuming. Through switching devices in landline telephone exchanges, mobile phone users can also access the global landline network, effectively bringing everyone within speaking distance.

The mobile phone industry owes its growth to information technology which is in fact central and pivotal in any mobile system. The following technologies are often associated with mobile phones.

1. **PCS:** Personal Communications Service (PCS) is a wireless phone service somewhat similar to a cellular telephone service but emphasizing on personal service and extended mobility. It is sometimes referred to as digital cellular (although cellular systems can also be digital). Like cellular, PCS is for mobile users and requires a number of antennas to blanket an area of coverage. As a user moves around, the user's phone signal is picked up by the nearest antenna and then forwarded to a base station that connects to the wired network.
2. **TDMA:** Time Division Multiple Access (TDNA) is a technology used in digital cellular telephone communication that divides each cellular channel into three time slots in order to increase the amount of data that can be carried.

3. **CDMA:** It employs Analog-To-Digital conversion (ADC) in combination with spread spectrum technology. Audio input is first digitized into binary elements. The frequency of the transmitted signal is then made to vary according to a defined pattern (code), so that it can be intercepted only by a receiver whose frequency response is programmed with the same code, and so it follows the transmitter frequency exactly. There are trillions of possible frequency-sequencing codes. This enhances privacy and makes cloning difficult.
4. **GSM:** Global System for Mobile (GSM) communication is a digital mobile telephone system that is widely used in Europe and other parts of the world. GSM uses a variation of (TDMA) and is the most widely used of the three digital wireless telephone technologies (TDMA, GSM and CDMA). GSM digitizes and compresses data, then sends it down a channel with two other streams of user data, each in its own time slot. GSM is in fact, the de facto wireless telephone standard in Europe.

Today, mobile phones are proliferating as handsets are getting cheaper and call rates are declining, bringing them within the reach of the common man. They provide an array of functions (some very simple and others very sophisticated). Some of the popular functions which are based upon IT are as follows:

- **SMS (Short Messaging Service):** Small text messages can be exchanged between people who do not believe in long verbal conversations over communication channels. In fact, today SMS has gained popularity as a medium for exchanging messages.
- **Address Book:** It is a store of contact information maintained on the mobile handset or the central server. It does away with the usual problem of maintaining an usual address book and allows the phone user to dial numbers without having to bother about carrying a bulky file-o-fax or telephone diary along.
- **Schedules or To-Do Lists:** You can store a list of important tasks that you wish to accomplish. Most mobile phone software also provide options for appointments and reminders associated with these tasks.
- **Send or Receive E-Mail:** Thanks to WAP technology, it is now possible to access your e-mails by using your mobile phone. Popular portals like Yahoo and Rediff offer a facility whereby the users get automatic alerts on their mobile phones as soon as any new mail arrives. You can also use your mobile phone for chatting by using your MSN or ICQ account.
- **Get Information Updates:** All mobile service providers now provide add-on facilities for their subscribers to receive regular updates on news, entertainment, stock market prices. This is done by integrating Web-based databases with the mobile users' database. Service providers also use this ability to advertise for new products, services and schemes.

As you can appreciate, all the above facilities are based upon the usage of electronic databases and intelligent software available on the mobile phone. Due to the global trend of convergence the dividing line between information technology and telecommunications technology is getting increasingly blurred. Today's computers combine phone, fax, television, VCD/DVD drives, stereo—all in one seamless bundle.

NOTES

Address book: It is a store of contact information maintained on the mobile handset or the central server

NOTES

Hotel Management

The hotel industry is an integral part of the tourism industry, which is a vital source of revenue and foreign exchange for a country's economy. A vibrant hotel industry means greater employment generation. However, since this industry relies on an easy and quick availability of information, the role of IT in its development and growth cannot be over-stressed. In fact, IT has revolutionized the hotel and tourism industry. Due to IT, you can have all the information about the tourist spots, hotel infrastructure, room availability, tariff details, online booking, etc. at the click of a button. IT plays a critical role in improving the hotel's performance because of its potential of creating customer relationships and the flow of information between the people and customers.

There are numerous instances of use of IT in hotel industry. Some of these are as follows:

- Today's hotel management software means that the moment a guest expresses any interest in staying at the hotel, till the time he checks out, all transactions with him (room charges, food and laundry bills, business centre, heath centre, hiring cars, etc.) are recorded electronically, making information available at the click of a button.
- Many leading hotels offer online booking facility for tourists and guests. This makes it very easy for the tourist as he has beforehand knowledge of room availability and charges. There are several Websites totally devoted to this. A tourist, for example, can specify the city and his budget. Based on this information, the search facility throws up a complete list of hotels available. Moreover, the tourist can even specify his preferred location. Once the hotel is identified, booking can be made online by using an internationally valid credit card.
- Nowadays, most of the hotels have computerized their records. As a result, it is very easy to know the details of room availabile at a particular time. The information about the occupant is also available instantly. This computerized system typically integrates all hotel MIS functions into one system. Inter-continental hotels and rsesorts use a global strategic marketing database. All these are examples of use of IT in hotel industry, which have made significantly transformed operations and profitability.
- Hotel information systems help users to access guest database information and use the information to create attractive one-to-one reservation confirmations, e-mail marketing and sales messages, custom reports and e-mail comment cards to reinforce guest relationships.
- Information technology is being increasingly used by International hotel chains to formulate and align their corporate strategies.

Web Publishing

Traditionally, 'publishing' has meant dealing with printers, paper, distribution, expensive infrastructure and static content. The drawbacks of traditional publishing are that they require a huge amount of investment, the productivity is low as a lot of manual and machinery work is involved, the content published cannot be changed easily and the scope of marketing the product is very limited. All these drawbacks have been overcome by the development of Web publishing.

Web publishing is an umbrella term for putting content on the World Wide Web and includes all support arrangements required for it. It includes custom Web designs for Web development, Website hosting and e-commerce. Originally, Web publishing simply meant putting selected content on paper into HTML over a Website for public access. This is also known as **ipaper**. This method of publishing is not widely used anymore as professional web publishers now use modern software, such as content management systems for rearranging the structure of a Website and making its content dynamically modifiable.

NOTES

Ipaper: Putting selected content on paper into HTML over a Website for public access

The most important tool of information technology used in the process of Web publishing is the World Wide Web. This makes content available twenty-four hours a day, seven days a week, to anybody in the world who is connected to the Internet. The only requirement for publishing and viewing the content online is a computer or a handheld device which has an Internet connection and a Web browser. The scope of Web publishing in terms of penetration is very high with an estimated 1.5 billion Internet users worldwide, as of 2007. The relative low cost of buying a domain name and hosting a Website is another major driver behind the large amount of online data available over Websites.

Financial Accounting

Financial accounting was one of the first business functions for which software applications were developed. The importance of financial accounting and management for any business cannot be overemphasized, but the scale of transactions, the repetitive and structured nature of the data and the sheer volumes involved in the case of large corporates makes an ideal case for computerization. Computerizing accounts also takes the drudgery out of bookkeeping, which means that accountants can now concentrate more on analyzing information rather than on devoting countless hours merely in filling out vouchers and updating registers and ledgers.

Typically, this is how a computerized accounting system works—the accounting clerk makes the voucher directly on the computer by using a financial accounting software package. The voucher on the screen looks very similar to a regular paper voucher and is in fact much simpler to fill because things like current date and voucher number are generated automatically. The appropriate account names that have to be debited or credited need not be typed but simply selected by the click of a mouse from a list of all ledger accounts. Appropriate checks and validations are also built into the accounting software which reduces the chances of errors. Unless, for example, the total of all debit accounts equals to the total of all credit accounts, the software will not allow the voucher to be saved.

Once the basic data has been entered into the computer voucher, the accountant can print out as many copies as required. Unlike a manual accounting system where the voucher, once prepared, has to be entered into the daybook and then posted in the relevant ledger account, the computer software does this automatically. In fact, the moment the voucher is entered and saved it is not only automatically posted to all the relevant daybooks and ledger but also up-to-date trial balance, profit and loss account and balance sheet can be generated instantly showing the downstream effects on each one of them. Since, there is no time lag between voucher preparation and posting, the accounting software always shows up-to-date statement and final account.

NOTES

Depending upon the size of the organization and the complexity of its operation, different software packages are readily available in the market. At the bottom end are popular and inexpensive software such as Tally and EX which are quite sufficient for most small and medium scale-organizations. Tally provides an excellent user-friendly interface through which all the accounting transactions can be entered or modified easily and the user can see the effects of each transaction in all financial statements.

At the top end of the market is ERP (Enterprise Resource Planning) software like Oracle Financials, Baan, SAP, etc. which caters to the financial accounting and management needs of huge multi-location, multi-currency, multi-operations organizations like Nestlé, Pepsi, Coca Cola, Procter & Gamble, etc. Such a software is called ERP software, since it provides completely integrated solutions for all functions of a business like financial accounting, inventory, payroll, production planning and control, etc. Despite the fact that ERP solutions typically cost millions of rupees and are relatively much more complex to implement, they provide an excellent platform for ensuring that the company's system and procedure are uniformly followed across multiple locations (or even countries). Such systems also make it very easy to consolidate huge amounts of information from different profit centres and locations.Thus, effective, near real-time management information can be generated to assist apex level decision making.

Weather Forecasting

Predicting the condition or state of the atmosphere after a period of time and over a certain region(s) is known as weather forecasting. The professionals involved in the study and prediction of weather are called meteorologists. The state of the atmosphere is governed by various factors, such as temperature, humidity, wind speed, etc.

A few decades ago, man depended on the close observation of natural phenomenon and changes in atmosphere such as cloud formation, sky colour, wind speed, temperature, animal and insect behaviour to make weather predictions. Human senses and knowledge used to be the main driving factor behind these early predictions, which were limited to short-term forecasting and had low accuracy levels.

With the development in information technology weather forecasting has become a science rather than an art. Weather forecasting requires processing and analysing of huge amount of data very quickly. This makes it an ideal field for the application of information technology. The volume of data to be processed and the complexity of calculations that must be made in order to forecast weather with a certain degree of accuracy can be gauged by the fact that this task can only be performed by super-computers which work at phenomenally high speeds and can crunch huge amount of data very quickly.

The software and hardware tools provided by IT help in making accurate weather forecasts over longer time intervals. Large amounts of data are collected by weather balloons, satellites, sensors and radar instruments and fed into computers with huge processing power and data storage for quantitative analysis and weather modelling. Accurate assessments of the condition of weather over a period of three to six days can be made by using hydrological forecasts and warnings of extreme events can be issued over a lead time of five to ten days.

There is still a huge scope of development in the field of weather forecasting and information technology is driving it by developing better software for computer

modelling, building and designing weather monitoring sensors for data collection, analysis and growing channels of weather forecasting services and making systems with huge computing power and storage space available.

Remote Sensing

The retrieval of data and information regarding an object or phenomena without coming into physical contact with it is known as **remote sensing**. The devices used for recording such data are known as sensors and depending upon the method of retrieval there can be either recording or real-time sensors. The technique of remote sensing determines if it is active remote sensing or passive remote sensing.

In the active remote sensing, artificial radiation is bombarded over a particular region of interest and the reflected rays are detected by the sensors to collect data and relevant information. An example of active remote sensing is the radar technology.

Passive remote sensing only detects natural radiations of an object or the one reflected from its surrounding area. The remote sensors do not emit radiation for measuring values of the object. A remote camera setup to observe wildlife and natural phenomenon is a good example of passive remote sensing.

In earlier times, our forefathers used to find high ground or climb treetops to map the surrounding landscapes for information. Later, in the year 1858, balloonist G. Tournachon took photographs of Paris from his hot air balloon. Then, with the help of IT remote sensors, computer systems and software were developed to precisely monitor and collect geographic or spatially-referenced data. The preceding traditional drawbacks have been successfully overcome with the help of IT.

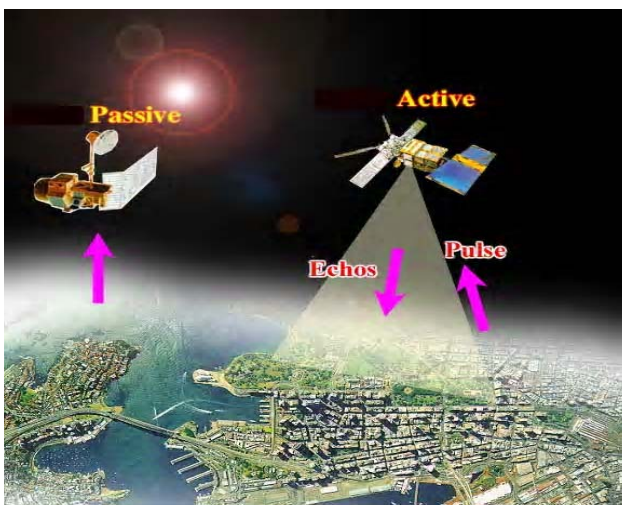


Fig. 5.12 *Remote Sensing*

The various applications of IT in the field of remote sensing are as follows:

(i) Software: Embedded software are used to process data from remote sensors and turn it into relevant information. They also control the functions of a remote sensor by judging the data returned from it. Image enhancement and grouping applications, help in clearing the interference from raw images (captured images from camera with minimally-processed data and huge detail) and can be used to transform multiple images into one high resolution continuous image.

NOTES

Remote Sensing: The retrieval of data and information regarding an object or phenomena without coming into physical contact with it

NOTES

Planning: Describe the formal procedure used in such an endeavour, like creation of documents, diagrams or meetings to discuss the important issues to be addressed, the objectives to be met and the strategy to be followed

(ii) Hardware: IT helps in designing customized hardware components for the purpose of remote sensing. The capabilities of a sensor can be optimized if they are redesigned for each application.

(iii) Telecommunication: Advancements in the communication between the sensor and the base station have helped in increasing the remote distance. Global environmental mapping would not have been possible without worldwide telecommunication. Figure 5.12 shows how remote sensing operates with the help of a satellite.

Planning

Planning in organizations—public and private—concerns both the organizational process of creating and maintaining a plan and the psychological process of thinking about the activities required to create a desired future on some scale. As such, it is a fundamental property of intelligent behaviour. The thought process is essential to the creation and refinement of a plan, or integration of it with other plans, that is, it combines forecasting of developments with the preparation of scenarios of how to react to them.

The term **'planning'**, is also used to describe the formal procedure used in such an endeavour, like creation of documents, diagrams or meetings to discuss the important issues to be addressed, the objectives to be met and the strategy to be followed.

Planning is a crucial aspect of an individual, organization and economy. It is done to attain growth, development and competitive advantage in a firm. Information technology tools have been a growing contributor to planning over a number of years.

It is a commonly acknowledged fact that with the right knowledge at the right time a firm can become the market leader of its products and services and continue to make profits for further growth. Therefore, planning helps an organization in facing and beating the competition. Second, the daily operations of an organization are becoming increasingly dependent on telecommunication and distributed networking processes.

Information and Communications Technology (ICT) tools greatly assist the planning process since they allow large amount of historical data to be processed and analyzed which form the major requirement for the future planning process. Also by using sophisticated scenario analysis tools, decision support systems allow the managers to know the repercussions of making long-term or policy decisions such as entering a new market, or introducing a new product or increasing the prices of goods and services being offered. These packages, by using a combination of complex algorithms, mathematical calculations, statistical analysis, etc. allow the managers to predict the outcome of such policy changes and therefore enable them to plan better.

So, whether it a case of a small grocery store deciding what to order (from its suppliers) for the coming week's sales, or a large multinational working in many countries trying to do the inventory forecasting for its thousands of stores, IT tools can be used to automate the basic number crunching (data collation and compilation) and make better decisions regarding the future.

NOTES

Applications in Medical science

Medical science is a branch of science that treats injuries and prevents and cures diseases by prescribing medicines or boosting the immune system of patients. IT has completely transformed the way modern medical systems work—from storing information about a patient's history to developing new ways of diagnosing patients and educating students in medicine (see Figure 5.13). IT has become such an integral part of the modern medical system that nowadays it is inconceivable to think how this industry worked without the aid of ICT.



Fig. 5.13 *IT Applications in Various Spheres of Medicine*

Developments in medicine due to IT have offered significant benefits to patients and healthcare systems. Research in hi-tech medicine, such as genetic research, DNA modification, hospital infrastructure, rapid ambulance services, etc., have been facilitated by IT. Medical scientists can now use computers to check the effectiveness of a drug against a disease by modelling their genetic structure on computer- based software and using high-speed processors to simulate the process.

The storage and rapid access to electronic medical records and its instant transmission over the Internet in large amounts is called teleconsulting where practitioners share patients' data across the world to diagnose patients cooperatively without experiencing their medical history. Videoconferencing between surgeons allows the sharing of expertise so that complicated procedures can be carried out by sharing knowledge in real time. This allows doctors to develop expertise without the need for supervising surgeons to travel. Operations can be performed in areas in which they would not ordinarily be accessible, potentially saving or improving many lives, with the help of IT.

Medical images are sometimes so complicated that they cannot be effectively analysed without using computers. They can not only improve the image quality, but also adapt images to fit in accordance to the doctor's wish.

Entertainment

No matter what your business is, IT can transform how you do business and bring many possibilities to life, IT has brought many changes in the entertainment field like video games, special effect movies, etc.

NOTES

1. **Video Games:** Games have been one of the most popular uses of computers. In fact, organizations like Attari, Nintendo and Sony who were developers of video games have been instrumental in the improvements in the multimedia capabilities of desktop computers. Till about a decade ago, when personal computers had severe limitations of disk storage, processing speed and memory size, only very simple uni-dimensional video games were possible. But with the development of much faster Pentium series of CPUs with in-built multimedia capabilities coupled with improvement in digital storage and acoustics, today's games are limited only by their creators' imagination and not by any technological hindrances. Today's games like Doom, Pokemon, PlayStation, Galaxian, Defender, etc. use very sophisticated graphics and sound techniques to create three-dimensional games.

   Some of the interesting developments in this area are as follows:
   - Virtual reality
   - Improvements in the specialized input devices like joysticks
   - Special game cards and enhanced graphic capability of CPU
   - Web games (Casinos)
2. **Special Effects in Movies:** Special effects in movies have come a long way since the early twentieth century. During the early years of movie-making, special effects were limited to time-lapse cinematography where hand-controlled dummies were brought to life by stop motion filming which meant manually moving the animated model a fraction of an inch and taking a snapshot.

   The early animation movies (popularly called cartoon films) involved a team of artists and painters who would painstakingly draw and paint each sketch frame by frame. The photography team would then click shots of these sketches at the rate of twenty four frames every second and then edit them into a story.

   Some of the interesting techniques used for creating special effects are as follows:

   (i) Digital Vompositing: Typically done through a process called 'Bluescreen' where the actors perform the scene in a studio in front of a large blue screen. A separate team of computer designers and artists create a virtual background (by mixing multiple photographs and computer-generated images). Later, the actors' footage is superimposed on the top of the background to create a seamless 'composite picture.'

   (ii) Time Slicing: In this technique a series of cameras are placed around the object of concern. All these cameras shoot pictures at precisely the same time. When these pictures are played together it appears as if there is one camera moving around the object. Coupled with other special tricks (such as slow motion photography as used in the Matrix series) this creates an ethereal effect.

   (iii) Computer-Generated Imaging (CGI): CGI techniques are used for creating scenes which are either not possible in real life or would be too expensive or dangerous to film.

   None of the above developments would have been possible without the fantastic developments in IT.

NOTES

With the arrival of the CD-ROM and the Internet, the entertainment industry made a huge leap into a new era with a winning card–multimedia. Armed with animated images, sounds, full-motion video and interactive capability, multimedia became a dominant factor in today's information age. The fast but steady growth of electronic technology allowed multimedia to gain popularity within a short span of time. Some applications of multimedia in the entertainment field are as follows:

- Games are the first thing that come to mind when we talk about multimedia. Multimedia capabilities are used to develop interactive games with sophisticated animations, 3D and sound effects. These games can be played on the computer, mobile devices or on the Internet. Live internet pay-for-play gaming with multiple players has attained significant popularity.
- Movies and cartoons unleashing full effects of multimedia, which were only available on VHS tapes, are now stored on CD-ROM (VCD) to allow the users to watch on their computer screen. Multi-layered digital versatile disk with more storage capacity and even higher processing speed, is making its way into the market and slowly replacing the CD-ROM. It can be used to view movies and play audio files. Multimedia can be used in voice mail, chatting and video conferencing as well. You can also do real time video conferencing with your colleagues spread across the globe.
- Another common application of multimedia is the advent of animated e-greeting cards for different occasions.
- Wedding albums and family histories can be created on the World Wide Web using the power of multimedia.
- Multimedia has also found its application in hotels, pubs, shopping malls, museums, cinema halls, where stand-alone terminals or kiosks are made available for guiding users. Printers are also usually attached so that users can walk way with a printed copy of the desired information.

3. **Multimedia in Marketing:** Advertising has become very prevalent in our daily surroundings, so for a product to stand apart, it is very essential to present it in a dynamic, visually stimulating manner to grab the attention of consumers. The business world is slowly rejecting run of the mill traditional methods (like placing ads in yellow pages, distributing pamphlets, etc.) and adopting solutions from the electronic era. Only companies with a nerve to radically change their marketing strategies for the new millennium will survive and be able to cater to the ever-changing customer's mindset. Applications of multimedia in the marketing field include:

   (i) **Presentations**: For launching the products of a company. Reaching the target audience with necessary technical services or products requires clear communication, stating the benefits and features, outlining its applications and any other product related details, all presented in a well-designed and interactive manner so that the users familiarise faster. Multimedia presentations are an excellent way to motivate, inform, and captivate a wide range of audiences via PC's, laptops, plasma screens, or kiosks delivered via CD-ROM or the internet.

NOTES

(ii) **Multimedia:** To create interactive product catalogues, training tutorials, buyer guides, and information directories with adequate search and navigation facilities to guide the user to easily trace the desired information. A buyer guide can list the nearby dealers, a comparison of the top brands, maps of the city and other helpful guest services.

(iii) **E-mail advertising or placing banner ads on the Internet:** An extremely cost effective method of launching a product, promoting an event or selling services. Effective use of multimedia in advertising can make potential clients sit up and make notice.

(iv) **Interactive applications:** A great way to build brand loyalty and drive inquiries or sales. Sales may be increased by allowing users to view product options real-time. Brand loyalty can be built by giving users a custom application that entertains, informs or assists them.

(v) **Motion, graphical elements, animation, audio and video:** Can be used to more effectively deliver sales, instructional or marketing messages to differentiate your firm from competitors.

Manufacturing

For any manufacturing firm, managing inventory is crucial. High inventory results in money being locked up unnecessarily, thereby reducing liquidity and indirectly profitability (if you offer immediate payment, most suppliers would be willing to offer you better rates). On the other hand, lower inventory of finished goods may lead to loss in sales or lower inventory of raw material may lead to disruption in production line. Optimum stock levels optimizes operational efficiency.

Most large manufacturing units typically need hundreds (if not thousands) of raw material components and produce many products. Managing optimal inventory of such large number of items is a difficult task. It is here that information technology again plays a very useful role. Inventory management software provides facility for specifying (and determining) the maximum, minimum and reorder levels for each item, so that appropriate levels of inventory can be maintained keeping in mind lead times and Just-In-Time (JIT) systems (if any) for component suppliers.

Basically, this is how a typical computerized inventory system works–a list of all the inventory items is prepared along with the maximum, minimum, reorder and current levels (quantity in hand as on a fixed date) for each item. This list is fed into the inventory software. Thereafter, all incomings (materials purchased or produced) and outgoings (sales or issues to production floor) are recorded through the inventory package. Since, the computer knows all the ins and outs for each item, it can track the exact quantity in hand for each. The package also generates reports for all the fresh stocks that need to be procured (based upon the levels specified). A variety of other useful MIS reports like aging analysis, goods movement analysis, slow and fast-moving stock report, valuation report, etc. can also be generated which assists the storekeeper and the accountants.

Some of the more sophisticated inventory packages (or inventory modules of ERP packages like Oracle financials, Baan, SAP etc.) automatically generate purchase orders (as soon as minimum level of any item is reached), provide automatic posting of

accounting entries (as soon as any purchase or sale is carried out) and generate analytical reports which show the previous and future trends in inventory consumption.

Some interesting innovations in usage of IT for better inventory management are as follows:

1. **Use of Barcoding System:** Bar coding is a technique which allows the data to be encoded in the form of a series of parallel and adjacent bars and spaces which represents a string of characters. A bar code printer encodes any data into these spaces and bars and then it is used to decode the bar codes by scanning a light source across it and measuring the intensity of light reflected back by the white spaces. Bar coding provides an excellent and fast method for identifying items, their batch numbers, expiry dates, etc. without having to manually type or read the data (see Figure 5.14).



Fig. 5.14 *Bar Code*

2. **Use of Hand Held Terminals (HHTs):** HHTs are simple devices that are used to communicate with any type of microprocessor-based device. The standard input device is the keyboard (typically more akin to the calculator, rather than the computer keyboard) and a small LCD display for the output. HHTs are compact, simple and rugged devices designed for the outdoor applications like collecting information about inventory from large warehouses, recording movement of goods in and out, etc.
3. **Internet and Intranets:** Many organizations (specially those following 'just-in-time' techniques) now have a system whereby the moment they receive an order or a request for an item (which is not in stock or whose stock is low), the inventory package automatically generates a purchase or supply order electronically and mails it to the preferred supplier—all this happens without any human intervention!

Business

Like banking, the insurance sector also deals contend with a lot of routine paperwork insurance policies, claims filed, survey or investigation reports, payment receipts, etc. IT provides a perfect opportunity to reduce costs and processing times.

According to the Insurance Journal, 'eighty-eight per cent of the insurers think that IT will become more important in driving efficiencies and cost-reductions in future. According to a new research released by RebusiS–an insurance technology solutions provider a further fifty-five per cent of respondents argued that IT is currently playing an 'important' role in driving efficiencies and cost-reductions, with 43 per cent contesting that IT is 'essential' to business efficiency.'

Typically, insurance companies use computerized databases to keep track of all the insurance policies, generating premium due statements, premium received receipts, lodging claims for insurance recovery, etc. Basically, all kind of transactions are recorded

NOTES

and processed through computerized systems. This not only enables insurance companies to provide quicker and more efficient service to their clients, it also allows them to minimize their risks and maximize their profits by enabling complex financial, economic and demographic analyses of their customers. By using sophisticated computer programs, an insurance company can determine which customer segments are growing the fastest, which are most profitable and which are riskier than others.

Although a lot of processes have been automated, things like insurance claims are still filed on paper forms first. The volumes involved are quite intimidating prompting some insurance companies in the US and Europe to outsource the entire data entry process to specialized offshore firms—many of such firms are in India.

This is how the typical process works–an insurance agent or the insured party fills up a paper form somewhere in the US. These forms are collected from multiple locations at one location such as the insurance company's head office. A team of professionals from the data entry agency (working in the insurance company's head office) scans these forms through high-speed scanners, generates image files for all the forms and then at the end of the day, using the Internet, transmits all these images to their data processing facility. Due to time difference, by the time this transmission is done at the end of the day in the US, it is morning in India. A team of trained data entry operators, using specialized software, views these forms (as images) on one portion of their screen and then types the same data in a database. Once the data has been properly verified and validated, the database is then uploaded on the network within a few hours. This means that the images that were sent from the US the previous night could be available in the US the next morning in the form of a computerized database. Of course, other than the effective use of IT, the time difference between the US and India has helped tremendously to make this 'zero time lag' system a great success. This system of outsourcing one of the business functions is called 'BPO' (business process outsourcing).

Another more sophisticated alternative to this is OCR—Optical Character Recognition—where the images are run through OCR software that automatically converts these into text. OCR is only feasible where the text quality is very (typically typed or computer printed matter) high. Since, OCR operations still produce only 90–95 per cent accurate text, human intervention is still required to correct the mistakes made by OCR systems. In the course of time, however, technological advancements can bring 100 per cent reliability and further change the face of remote-processing arrangements.

Education

Teaching has traditionally been associated with classroom instructions on a blackboard with the instructor (teacher) dependent almost entirely on his/her oratory and presentation skills for holding the attention of the class. From a student's perspective, he had to keep pace with the instructor's pace, which meant that the slower (though not necessarily less intelligent) student was at a natural disadvantage. Similarly some students were more interested in a more in-depth study than the others. Since, access to information was neither easy nor inexpensive, these variables had always posed a major barrier to learning.

Ever since the advent of information technology, the scenario has changed dramatically. Today, the instructor has a repertoire of information technologies. To make the lecture not only more interesting but also more informative, there are advanced electronic teaching tools available. These vary from simple slide presentations to full-blown multimedia presentations which have video clippings, sound effects, animation and graphics to explain even the most abstruse subjects in a simple and easy-to-understand manner. As an example, a medical student does not have to pore over boring textbooks to understand, human anatomy. Simple computer packages like 'Body Works' are available which explain the same by using photographs, images and graphics that make in-depth learning fun rather than a bore. Moreover, learning is not only faster but is retained longer when test is supported by visuals and sound clips. Multimedia has transformed both classroom as well as online (distance) and packaged (CDs, VCDs, DVDs, etc.) education, in terms of both content as well as interactivity.

Some of the interesting developments in IT for the education sector can be seen in the following:

1. **Computer-Based Training (CBT):** In most of the progressive institutes today, classroom sessions are complemented by CBTs. CBT typically comprise of user-friendly software in which the course syllabi is broken up into a series of interactive sessions. These sessions involve imparting a slice of knowledge to the student and then quizzing him to reinforce his understanding. Students have the option of going through these sessions at a time most convenient to them and a pace best suited to them. CBTs also provide an excellent medium for the student to learn by exploration and discovery rather than by rote. However, education software is often positioned as 'enriching' the learning process and not as a potential substitute for traditional teacher-based methods.
2. **Internet:** Thanks to the Internet, any and every type of information is available at the click of a mouse. Students no longer need to trudge long distances to visit a library and spend valuable time plodding through library catalogues to find the right information. By using a search engine, one can easily access the desired information. Also, knowledge is no longer restricted within the academic fraternity alone. Thanks to our networked world (intranet/internet) information dissemination is faster and widespread.
3. **Distance Learning:** Information technology has also made distance learning a reality. You need not be physically present in a Business School to do a management course from there. By innovative use of information technology, educational institutes have reached out to students who would otherwise never have been able to enroll with them.
4. **Computerization of Administrative Tasks:** Most academic institutes use computerized systems for student enrollment, fee management, examination, administration, etc. Enrollment forms, for instance, are now available on institutional websites, and examination results are usually available on the Internet. Some schools have also started collecting fees through the Internet by using credit cards.

NOTES

Distance Learning: Information technology has also made distance learning a reality

Check Your Progress

11. Fill in the blanks with appropriate words.
    a. A __________ is a peripheral device used to store and collect information.
    b. An ________ is a type of storage medium that stores the content in digital form which is written and read by a low intensity laser.
    c. The ________ reads data from the reflective surface of an optical disc by measuring surface changes in height and depth.
    d. ______ is intended to support continuous media intensive applications, such as personal video recorders, video JukeBoxes and Video-on-Demand (VoD).
12. State whether the following statements are true or false.
    a. The CD-ROM Mode 2 is used for compressed audio/video information.
    b. CD-R allows data to be erased and rewritten.
    c. DVDs are of the same dimensions as CDs, but store more than 6 times as much data.
    d. HHTs are simple devices that are used to communicate with any type of microprocessor-based device.

5.6 SUMMARY

NOTES

In this unit, you have learnt that:

- The compact disc is a thin, round plastic platter which is 12 cms in diameter and approximately one mm thick, with a hole in the center for a spindle.
- A polycarbonate layer of the CD that has the data impressed onto it, is coated with a mirror like metal film (aluminum or gold).
- Side of the CD that reflects light is available for use, while the opposite side is a silk-screen with the disc's identifying label or logo printed on it.
- One major limitation of CD-ROM is that they cannot be used to store data, but only to read data that was stored on them by the manufacturer. However, there are recordable CD's also known as Compact Disc-Recordable (CD-R).
- Another type of CD is CD-RW (Compact Disc-Rewritable) which not only allows data to be written but also allows erasing, thereby making the CDs reusable.
- A disk drive is a peripheral device used to store and collect information.
- DVD is also an optical disc storage media format. Its main uses are video and data storage. DVDs are of the same dimensions as compact discs (CDs), but store more than six times as much data.
- An optical drive is a type of storage medium that stores the content in digital form which is written and read by a low intensity laser. The laser reads data from the reflective surface of an optical disc by measuring surface changes in height and depth.
- Recent advances in compression, storage and communication technologies have resulted in the creation of applications that involve storing and retrieving multiple data types, such as text, audio, video, imagery, etc., collectively referred to as multimedia. These applications require the development of file systems that can efficiently manage the storage and retrieval of multiple data types referred to as integrated multimedia file systems.
- The Multimedia File System (MMFS) was specifically designed to provide a high performance network interface for storing and retrieving multimedia data.
- The MMFS or Multimedia File System maintains association between related files and also helps in storing multimedia data types, such as MIDI files, still images and video animation frames with a universal standard.
- MMFS is intended to support continuous media intensive applications, such as personal video recorders, video JukeBoxes and Video-on-Demand (VoD).
- In the twenty-first century, IT provides many services like airlines, hotel management, Web publishing etc. Some of these services are explained.
- Web publishing is an umbrella term for putting content on the World Wide Web and includes all support arrangements required for it.

NOTES

- The retrieval of data and information regarding an object or phenomena without coming into physical contact with it is known as remote sensing.
- Bar coding is a technique which allows the data to be encoded in the form of a series of parallel and adjacent bars and spaces which represents a string of characters.
- Typically, insurance companies use computerized databases to keep track of all the insurance policies, generating premium due statements, premium received receipts, lodging claims for insurance recovery, etc.
- Advertising has become very prevalent in our daily surroundings, so for a product to stand apart, it is very essential to present it in a dynamic, visually stimulating manner to grab the attention of consumers.

5.7 ANSWERS TO 'CHECK YOUR PROGRESS'

1. CD is a thin, round plastic platter which is 12 cms in diameter and approximately one mm thick, with a hole in the center for a spindle.
2. The CD recording method makes use of optical recording – using a beam of light from a miniature semiconductor laser.
3. CD-Rewritable (CD-RW) discs allow data to be erased and rewritten. So a CD-RW can be used repeatedly.
4. A disk drive is a peripheral device used to store and collect information.
5. An optical drive is a type of storage medium that stores the content in digital form which is written and read by a low intensity laser.
6. Some of the areas in airlines where IT has been successful are as follows:
   (a) Online ticket reservation.
   (b) Flight and seat availability.
   (c) Last minute ticket auction.
7. Infotel is used by Bharat Sanchar Nigam Limited for managing their telephone exchanges.
8. A cellular phone is primarily a radio. It has a base station that consists of a tower and a small building containing the radio equipment.
9. Web publishing refers to putting content on the World Wide Web. It includes custom Web designs for Web development, Website hosting and e-commerce. It is also known as ipaper.
10. Bar coding is a technique which allows the data to be encoded in the form of a series of parallel and adjacent bars and spaces which represents a string of characters.
11. (a) Disk drive, (b) Optical drive, (c) Laser, (d) MMFS
12. (a) True, (b) False, (c) True, (d) True

5.8 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What are different types of CDs?
2. What is the use of DVD-RW?
3. When the maximum speed is achieved in DVD-ROM?
4. How video conferencing is useful in medical science?
5. Who invented telephone service?
6. How information technology is used in hotel management?
7. What are the applications of multimedia in marketing?
8. Write the limitations of passive remote sensing.

Long-Answer Questions

1. How is information represented in multimedia applications? Explain with suitable illustrations.
2. Describe the recording process of CD-ROM?
3. Discuss the various formats of DVD and list their features.
4. What are the uses of computer in medical science?
5. Explain the technologies, associated with mobile phones.
6. What is remote sensing application? Explain briefly.
7. Discuss the uses of hand held terminals (HHTs).
8. Describe the importance of computer-based training (CBT).

UNIT 6 VIRTUAL REALITY

Structure

NOTES

6.0 INTRODUCTION

In this unit, you will learn about the basics of virtual reality. Virtual Reality (VR) is basically a way of simulating or replicating an environment and giving the user a sense of being present there, taking control and personally interacting with the environment with his/her own body began in military and university laboratories more than 30 years ago. It may be called artificial reality, cyberspace or synthetic reality.

You will learn about Virtual Reality Operating System (VROS). Virtual reality operating system consists of several software subsystems and requires extra time to reconfigure. It avoids hardwired configurations because a participant in the virtual world is free to engage in all the resources of the operating system. The Distributed Virtual Environment (DVE) system is an Internet based multi-user virtual reality system in which participants are able to navigate the 3-D virtual world and also able to work with various applications and other users.

You will also learn about virtual environment displays and orientation tracking. Virtual environment displays refer to a possibility of applying virtual reality technology. It highlights a synthetic construction using Virtual Reality Modeling Language (VRML) and appropriate simulation technique. The main characteristic of this environment is to include remote environments for the computer-supported cooperative work, entertainment and space exploration, scientific and architectural visualization. A Visually Coupled System (VCS) is defined as the subsystem of the multimedia environment, which was developed by Birt and Task in 1973. The three components of VCS are a head or helmet-mounted (head-directed) visual display, a means of tracking head of eye-painting director and a source of visual information.

Finally, you will learn about the intelligent Virtual Reality Software Systems (VRSS). It implements the notion of an intelligent virtual environment, in which alternative reality can be defined through a symbolic description of the virtual world's behaviour. This software architecture is based on an integration layer, which consists in an event-based system, relating the visualization engine to the behavioural layer.

6.1 UNIT OBJECTIVES

NOTES

After going through this unit, you will be able to:

- Discuss the significance of intelligent multimedia system
- Identify desktop virtual reality systems
- Explain what VROS is
- Discuss distributed virtual environment system
- Explain virtual environment displays and orientation tracking
- Define the basics of intelligent virtual reality software systems

6.2 INTELLIGENT MULTIMEDIA SYSTEM

Intelligent multimedia system provide an advanced technology for listening music and watching movies. This system allows users to avail the microphone facility as an analog input to sing and speak. That is why, it is not only a multimedia system but an intelligent multimedia system implementing various features to help the users to synthesize, modify and create multimedia records. Examples include adjusting the volume of music and voice. The voice can be echoed as it is generated in the theatre. Transferring the analog signal into digital format, the audio data is manipulated. The audio data is recorded in real time by passing to the system unit. For this, RS232 COM port is used to process the data smoothly. The kernel of the system is controlled by ATMega32 microcontroller. Basically, voice signal is an input by the microphone and music signal is produced by music application, for example, MP3 player on PC. Both audio signals pass through a preamplifier boosting to its signal strength. To transform analog signal to digital data, an analog to digital converter chip (ADC0801) is used. This data is propagated to the microcontroller. The whole technique allows transferring of the music data to the system unit using the intelligent multimedia system. At last, this signal is passed to the speaker for generating a wonderful melody. Java and Visual C++ in Windows programming provide effective features to the voice signal via the Graphical User Interface (GUI). The intelligent multimedia system allows users to effectively use the microphone by speaking and singing. It is the intelligent multimedia system in which various features are allowed to be modified by users. Users can synthesize, modify and create the multimedia record. Users can adjust the volume of music and voice. The sound is echoed with the help of intelligent multimedia and has the same effect as that of a theatre. For example, the music application, MP3 players is run so that digital to analog conversion is possible and vice-versa. For this, Windows XP is used to access the GUI features that can be controlled by the system by clicking the mouse. The mouse click is turned into a command to the microcontroller through the COM port. A Digital-To-Analog Chip (DAC0808) is used to transform the digital output to the analog output making the sound effect more dulcet. An integrated media editor contain specialized editing functions for each type of information, and allows the user to easily edit messages. The system allows users to create, store, retrieve, send, receive, sort, reply to, forward and delete messages. In case the user sends a message to other users whose system does not have multimedia capability, they can also run the multimedia system.

Intelligent multimedia systems consists of multimedia database, multimedia message, hypermedia and various virtual reality systems. It examines the potential survey of supporting hardware, such as Digital Video Interface (DVI) technology, Compact Disk Interactive (CD-I), Compact Disk Read-Only Memory Extended Architecture (CD-ROM XA), etc. to build multimedia computer systems. The multimedia presentation becomes intelligent by employing, powerful computer workstations equipped by high-resolution environment displays and high capacity magnetic and optical storage units. It provides functionality of hypertext by adding additional components, for example, two and three dimensional structured graphics, spreadsheets, paint graphics, sound, video and animation. Multimedia system becomes more intelligent if it is activated by node links, such as highlighted words, locations or graphic images (landmark view is symbolized by oil symbol ‘  ’), other types of markers, etc. This is called smart navigation. The graphic browsers present major nodes that allow visitors to move the nodes. It generates linear files to maintain multimedia database systems. The characteristics of intelligent multimedia system are as follows:

- It utilizes object oriented architecture.
- Its hypermedia fashion is linked to external objects, visual pages and audio annotations.
- It provides time driven multimedia objects, such as video sequence, shaded icons and full motion video information.
- It requires broadband optical fiber network having speeds of 2 to 5 gigabits per second.
- It meets the user’s need by built-in programming languages known as Slate Extension Language (SEL).
- It provides real-time conferencing, which is generally not a feature of regular multimedia system to transform the computing environment. It includes movement, tactile and orientation tracking system to enter a virtual reality.
- It is equipped with special earphones, gloves and body suit. If a user enters a color options the 3-dimensional virtual environment displays as a cartoon.
- It provides advanced virtual environment for multiple users. For example, NASA developed visualization for planetary exploration project that is visualized by virtual reality system allowing users to explore the planets of our solar system.
- It provides specialized intelligent system known as ‘intelligent agents’. These monitor multimedia databases to capture the relevant information, guiding the user in analysing retrieved information using textual, statistical and other analytical tools. It helps the user to create new intellectual works from original and retrieved information.
- It prefigures intelligent agent system to emulate voice-recognition system. Examples include, reminders, scheduling meetings, taking phone messages and checking plane reservation. A quintessential is an object lens system developed at the Massachusetts Institute of Technology which permits users to construct agents performing a variety of user-defined actions, such as combining e-mails, hypertexts, sorting and filing the incoming mails. It also

NOTES

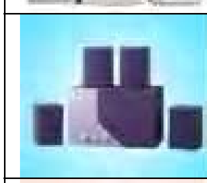
Intelligent multimedia systems: Consists of multimedia database, multimedia message, hypermedia and various virtual reality systems

allows users to generate new information requiring more capacious and powerful computer hardware and a security system to individual end-user accounts.

Table 6.1 shows the tools used in intelligent multimedia system.

NOTES

Table 6.1 *Tools used in Intelligent Multimedia System*

Tools	**Functions**
	This speaker accepts 80Hz-20KHz input sensitivity.
	This CD/CD-R/CD-RW can be implemented in PC.
	This audio CD Player PCD-901 plays CD-R/CD/DC-RW.
	This headphones driver keeps input power 50MW frequency response. This 20Hz - 20KHz sensitivity 98dB/MW at 1 KHz.
	Video VCD player is used to play the audio/video files.
	The wireless speaker has peak power maintaining system frequency response as 35Hz-20kHz degree of distortion.
	This multimedia speaker W-400D has 4.1 channel audio output super-woofer output with low noise Power audio as 560W (PMPO).
	In the USB portable panel speaker, no power needed. Its size is 150mm (W) x 120mm (H) Weight: 170.
	This PC cam CMOS sensor Image resolution uses 352 x 288 File format and its AVI Image catch speed is 15fps for processing the video files.

6.3 DESKTOP VIRTUAL REALITY

Virtual Reality (VR) refers to a technology that began in military and university laboratories more than 30 years ago. It may be called artificial reality, cyberspace or synthetic reality. It is basically a way of simulating or replicating an environment and giving the user a sense of being present there, taking control and personally interacting with the environment with his/her own body. In fact, VR has the ability to make the artificial as realistic and even more realistic, than the real.

The essence of VR lies the in computer-generated 3D world introduces and initiates. Its interface participants in a 3D synthesized environment generated by one or more computers. It allows them to act in real time within this environment by using one or more control devices and involving one or more of their physical senses.

NOTES

Virtual Reality (VR): It is basically a way of simulating or replicating an environment and giving the user a sense of being present there, taking control and personally interacting with the environment with his/her own body

6.3.1 VR Technology and Tools

VR originated in the second half of the 1960s. It comprised the Head-Mounted Display (HMD) as the first device that provided immersive experiences with computer-generated imagery. An HMD contains two small stereoscopic screens positioned just a few inches in front of the eyes; a motion tracker continuously lets an image generating computer adjust the scene to the user's current view. As the user moves his head, the images shift within the wide-angle field of vision to create an illusion of movement.

Immersive technologies can now include 3D head-gear with stereoscopic vision for look around and walk through, directional auditory input. They can also include voice recognition, data gloves, hand-held wands and other tactile or haptic tools for manipulation and control of virtual objects (see Figure 6.1). Even body suits wired with biosensors for advanced sensory input and feedback can be used. The data glove is a key interface device that uses position-tracking sensors and fiber-optic strands running down each finger. This allows the user to 'touch and feel' a virtual object. The user can pick up an object and do things with it just as he/she would do with a real object.



Fig. 6.1 *Head Gear and Data Gloves*

The Binocular Omni-Orientation Monitor (BOOM) system uses a screen and a stereo optical system housed in a box attached to a multilink arm. The user looks into the box through two holes, sees the virtual world, and controls action through sensors linking the arms and the box (see Figure 6.2).

NOTES



Fig. 6.2 BOOM

More successful and currently popular is the Cave Automatic Virtual Environment (CAVE), where the illusion of immersion is created by projecting stereo images on the walls and floor of a room-sized cube. Participants wearing lightweight stereo glasses enter and walk freely within the CAVE room (see Figure 6.3), while a head-tracking computer system continuously adjusts the stereo projection based on the current position of the viewer.

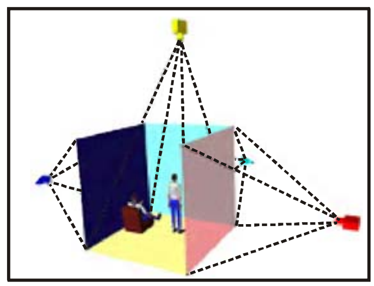


Fig. 6.3 CAVE Room

Applications

As the technologies of virtual reality evolve, the applications of VR become unlimited. The applications of VR include the following:

- Training in a variety of areas (military, medical, equipment operation, etc.).
- Education through virtual classrooms.
- Designing (virtual prototyping) of real and abstract systems.
- Architectural walk-through.
- Simulation of assembly sequences.
- Equipment stress testing and control.
- Accident investigation and analysis.
- Business and economic modelling.
- Entertainment.

Virtual Reality Operating System (VROS): It consists of several software subsystems avoids hardwired configurations because a participant in the virtual world is free to engage in all the resources of the operating system

6.4 VIRTUAL REALITY OS

Virtual Reality Operating System (VROS) consists of several software subsystems and requires extra time to reconfigure. It avoids hardwired configurations because a participant in the virtual world is free to engage in all the resources of the operating system. For this reason, it permits the world to respond immediately to the participant.

NOTES

It requires broad bandwidth and multi-sensory interaction with VR systems which creates a demand for sensor integration. This model corresponds to the VR system that helps multimedia applications to accommodate multiple concurrent users. This approach makes VROS essentially a distributed OS by providing the detailed specifications of system functionality, scalability of conceptual design, fault-tolerant shared memory and heterogeneous communication. The VROS provides an architectural framework that allows an extensible collection of interfaces, protocols and network security management systems. It provides the special devices which are commonly used in head-mounted display and VR applications. Main accessories provided by this operating system are spatial sound, data gloves, 3D position tracker, rapid interactive prototyping and digital libraries of virtual factory.

Features of VROS

The various features of VROS are as follows:

- It interconnects with 3D motion in tracking devices.
- It supports backup networking, for example, Jumbo frames and TCP offload.
- It also facilitates add/remove option for VHD and pass through disks.
- Its guests with high performance para-virtualized in guest drivers (up to 8-way virtual SMP and 255GB is for maximum guest memory).
- It builds up the solutions for improved disaster recovery and delivers a high availability of recovery of files.
- It provides compatibility for standard operating systems and encapsulates a complete computing environment.
- It also provides the mechanism of hardware independency by installing VGA card, physical motherboard, network card controller etc.
- If a virtual machine crashes, it avails the security applications run under a virtual environment.
- It provides portability of hardware resources, for example, virtual machine can be moved to other data storage medium (from pocket-size USB flash memory card to enterprise Storage Area Networks).
- This operating system synthesizes and integrates the sensor databases, interfaces, etc. The broad-bandwidth display and the multi-sensory interaction of VR systems create severe demands for sensor integration.
- It uses multi-user mainframes and provides virtual systems with a developer toolkit, position sensing, sound and graph programming system for third-party etc.
- Here, the operating system uses the client operating system and can not be accessed directly. It copies the multiple versions of other OS, for example, IBM's VM operating system, Windows OS, etc.
- It corresponds to mail servers, print servers, file servers, network servers, security servers, backup servers etc.

Characteristics of Immersive VR

Immersive virtual reality is a hypothetical future technology that exists today as virtual reality art projects. It casuists of immersion in an artificial environment where the user feels just as immesed as they usually feels in conserses reality.

NOTES

The unique characteristics of immersive virtual reality can be summarized as follows:

- Head-referenced viewing provides a natural interface for navigation in three-dimensional space and allows for look-around, walk-around and fly-through capabilities in virtual environments.
- Stereoscopic viewing enhances the perception of depth and the sense of space.
- The virtual world is presented in full scale and relates properly to the human size.
- Realistic interactions with virtual objects via data glove and similar devices allow for manipulation, operation and control of virtual worlds.
- The convincing illusion of being fully immersed in an artificial world can be enhanced by auditory, haptic and other non-visual technologies.
- Networked applications allow for shared virtual environments.

Shared Virtual Environments

By using a BOOM device, a CAVE system, or a Head-Mounted Display networked users at different locations anywhere in the world, can meet in the same virtual world. All users, are present in the same virtual environment from their respective points of vision. Each user is presented as a virtual human or avatar (see Figure 6.4) that is actually a computer generated character. Therefore, the users can interact not only with the virtual environment itself, but also with each other by using avatars.



Fig. 6.4 *Avatar*

Non-Immersive VR

Nowdays, the term 'Virtual Reality' is also used for PC-based applications that are not fully immersive. These applications are known as **Desktop VR**, it focusses on mouse, joystick and space/sensorball-controlled navigation through a realistic 3D environment, stereo viewing via stereo glasses, stereo projection systems, etc. using a graphics monitor under computer control. Desktop VR began in the entertainment industry, making its first appearance in video arcade games with extensive use of sophisticated computer graphics and animation technology.

QuickTime VR software packages, such as *PixMaker*, *PanaVue Image Assembler*, and *VRWorx* helps the user create desktop VR environments for a modest software purchase. The additional cost of a standard digital still camera needs to be paid for.

NOTES

Applications

As the technologies of virtual reality evolve, the applications of VR become literally unlimited.

Useful applications of VR include:

- Training in a variety of areas (military, medical, equipment operation, etc.).
- Education through virtual classrooms.
- Designing (virtual prototyping) of real and abstract systems.
- Architectural walk-through.
- Simulation of assembly sequences.
- Equipment stress testing and control.
- Accident investigation and analysis.
- Business and economic modelling.
- Entertainment.
- ***Telepresence Systems*** that permit operation and control of devices and processes while working at-distance. They can currently be seen in telemedicine, teleoperation of industrial equipment, and telerobotic control of engineering, manufacturing and other processes. Generally, telepresence systems are based on haptic or tactile input technology, immersing a participant in a real world captured by video cameras at a distant location and allowing remote manipulation of real objects via robot arms and manipulators.

VR is a very valuable instructional and practice alternative in situations where exploration of environments or interactions with objects or people is impossible or inconvenient like battle, firefighting, anti-terrorism training, nuclear decommissioning etc.

Flight Simulation is an example of VR application in pilot training. The simulator creates a virtual cockpit of an aeroplane with all relevant gadgets and switches. The front wind shield of the cockpit is replaced by an even shaped high resolution screen. The screen displays different image sequences, the same as what a pilot sees from the cockpit wind shield during plane landing, take off or glide. The trainee pilot has a sensation of flying a real aircraft and thus gets the requisite training.

VRML

VRML stands for *Virtual Reality Modelling Language*, which is a standard tool for the modelling of three-dimensional virtual environments that are functional and interactive. Just as HTML became the standard authoring tool for creating cross-platform text for the Web, so VRML developed as the standard programming language for creating Web-based VR. Like HTML VRML provides integrated hyperlinks in the Web. The viewing of VRML models via a VRML plug-in for Web browsers is usually done on a graphics monitor under mouse-control and, therefore, not fully immersive. The current version VRML 2.0 has become an international ISO/IEC standard under the name VRML97.

VRML: Stands for Virtual Reality Modelling Language, which is a standard tool for the modelling of three-dimensional virtual environments that are functional and interactive

NOTES

VR Related Technology

A related branch of VR used for similar purposes is called ***Augmented Reality*** *(AR)*. AR allow for the viewing of real environments with superimposed virtual objects.

A special headgear creates stereoscopic 3-D images of virtual objects (or computer-generated renditions of real objects that cannot be seen directly) and superimposes them on to the real world view. By tracking the exact position of the headgear, the virtual objects can be registered in the correct positions in the real world. By wearing a glove or other device that is also tracked, you can touch and interact with these virtual objects. The user interface to this technology will likely be voice, head and body gestures. Thus, AR supplements rather than replaces the user's real world.

6.5 DISTRIBUTED VIRTUAL ENVIRONMENT SYSTEM

The Distributed Virtual Environment (DVE) system is an Internet based multi-user virtual reality system in which participants are able to navigate the 3-D virtual world and also able to work with various applications and other users. Basically, this system is a software system through which people who are geographically dispersed across the world can interact with other by sharing a consistent environment in terms of space, presence and time. These environments usually aim for a sense of realism by incorporating the 3-D graphics. DVEs are widely used in virtual shopping mall, interactive e-learning, multiplayer online gaming, etc. In DVE system, each user is represented as an entity called 'avatar' whose state is controlled by the input commands. The basic idea behind distributed virtual environment system is to let the users immerse into virtual tour and also let them to work in virtual world to avail the required resources in distributed manner. The virtual reality data maintains the addresses of artificial world and spatial databases so that distributed virtual environment can structure and store the data and control to the artificial world. This artificial world is used by virtual reality navigators, Internet viewers and standalone hardware hubs and stations. User navigation and interaction uses existing datasets 'on-the-fly' theme, whereas the existing dataset manages the fixed sequences interactively with dynamic behaviour of objects. The scripts, such as timers, collisions, timers, are evaluated if the object is replicated. You can find the virtual environment effect in virtual movies, for example Shrek, Kung Fu Panda, etc, were given by the same production house. Shrek is a 3-D virtual movie and online streaming available for Web audiences, whereas Kung-Fu Panda is a 2008 American animated virtual movie was produced by DreamWorks Animation's studio in Glendale, California and distributed by Paramount Pictures. The virtual movie 'How to Train Your Dragon' is about a child named 'Hiccup', young dragon who settles down when he became the friends with young Dragon. This virtual movie is a 2010 computer-animated fantasy film by DreamWorks Animation loosely based on the 2003 book of the same title. The animated world was created in this movie and it brings extra ordinary landscape of virtual world along with full length animated feature. This movie is fledged by most of the multimedia concepts and focused with virtual environment system. The whole scenario presents the virtual environment system and makes feeling to the multimedia users. DVSE or Digital Virtual simulation Environment is considered as prototype. The characteristics of DVSE are as follows:

- This prototype supports the development of virtual environments, user interfaces and applications based on shared 3-D synthetic environments.
- It is tuned and focused to multi-user applications in which the network participants work across network.
- It is supported by WWW and HTTP/HTML/FTP/MIME compliant.
- It includes virtual battlefields, spatial models of interaction, virtual agents, real-world robot control and multi-modal interaction.
- It exports Virtual Reality Modelling Language (VRML) and other 3-D formats.

NOTES

Web represents WWW. It refers to the client-server service. The client uses the browser to access the information from the server. There are many sites that providing information that is available at various locations on the Internet. The Web represents a hypertext system which presents information from across the net. It was developed first time in the month of 1989 by Tim Berners Lee. The at the European Laboratory, CERN, Switzerland. Internet is considered to be a combination of portability, user-friendly features and flexibility of information. Web sites keep the important operations on the right side of the screen. This includes the heading of a section, phone, URL description and the domain details. A comprehensive user's guide must be provided to the user which would give relevant information to him pertaining to the Web site. This can be done after the Web site has been launched successfully. A Web site can be launched by using the Google and Yahoo local listing. They optimize the search engine facilities for your Web site that offer moderate list of options, searchable description and the third party data providers, such as Super Pages, Yellow Pages, City Searches, etc. They also offer a free service in which you can enter your Web site domain area, tell the customer Who You Are ('WYA') facility, get reviews, etc. There are many factors that decide the success of a Web site on the Web. The following factors must be kept in mind at the time of creating and launching a Web site:

- The Web site is launched for the site using a File Transfer Protocol program (FTP). It is an economical option. The owner of the Web site must instruct the Web designer and the system analysts to implement FTP for the Web site.
- The Web hosting firm provides space on the server to the Web site.
- The owner of the Web site removes the 'teething' problem before launching the Web site. Half-finished or incomplete Web site can discourage visitors from returning to the Web site. For example, if your site provides e-commerce services, you must intend to provide the services on time and the services should be of value to the visitors too. Teething problem refers to the initial issues related to the set up and layout of the Web screen.

The required devices for DVES are supported by computer, video capture device, video capture device, video editing software, DVD recording software, DVD burner (to record physically DVD) (see Figure 6.5 (a), (b), (c)).



(a) (b) (c)

Fig. 6.5 *(a) Record from Webcam (b) Record from External Device (c) Record Your Computer Screen*

NOTES

FTP or File Transfer Protocol refers to an application protocol to exchange files between computers across the Web. It is the simplest method of downloading and uploading a file from a server, for example, downloading documents or articles from a Website. FTP uses TCP/IP to transfer data. For this, an FTP server and an FTP client are required. FTP also works on a client/server principle wherein an FTP client program is used to make a request to an FTP server (files can be stored on computers referred to as FTP servers). FTP supports programs involved in TCP/IP. FTP is used as a command line interface. For example, a commercial program provides GUI features at Windows' DOS prompt. Using FTP, you can update files on a server. Your Web browser can also make FTP requests to download programs you request from your page. For this, you need to login on an FTP server. The files are easily available because of anonymous FTP.



Fig. 6.6 *Conference Room Supported by Distributed Virtual Environment*

Figure 6.6 explains a virtual environment conference with three people simultaneously connected to the DVSE system. The people, represented by embodiments, interact, talk, and move in a shared and distributed environment. It spans several projects and groups, including spatial interaction models, high-level programming, agents, robots etc.

The distributed virtual environment system is supported by VoIP technology frequently. Hypertext is the main concept that makes the World Wide Web more than just another message transfer system that supports DVSE. The prefix 'hyper' usually means 'above' or 'beyond' and thus hypertext is like text, but goes beyond it in terms of functionality. The extra information in a hypertext document is used to tell the computer program that displays the file to a user how to format it. These additional elements along with the content of a document are commonly referred to as marking up the document. WWW hypertext documents the use of Hypertext Markup Language (HTML). HTML documents are as ASCII text files are arranged using a special structure of HTML elements that defines the different parts of the document and how they should be displayed to the user. Each element is described using special text tags that define its characteristics. An HTML document contains two prime parts known as head and body. The head element contains meta data and the title of a Web browser, whereas body element contains the text or information that will be displayed on the screen. S/MIME functionality offers to sign up the encrypted messages. It examines the message formats and prepare the enveloped data (contains encrypted content and keys for recipients), signed data. IPSec provides security services in the layer of IP selection. It uses security services in the layer of IP selection. It uses two protocols

known as authentication protocol and authentication header. Encapsulating Security payload (ESP) provides confidentiality services for the data packets and data filtering. It supports security parameter index (32 bits), sequence number (32 bits), payload data (variable), padding (0-255 bytes), pad length (8 bits), authentication data (variable). WWW and distributed virtual environment system build an online presence to global customers providing security and stability about the products and payment mode. It is also known as digital Web marketing. In 1990's, a solid core of expertise is reached in the field of Internet business that leverages the Internet as a tool to achieve strategic business success. Business consulting includes the features, such as e-business consulting, e-marketing consulting, Website usability, the Internet trends and the traffic and Search Engine Optimization (SEO) through WWW. At this stage, users can search the information and use Web properly. The main attraction of WWW for the distributed virtual environment system is to develop containing graphics and audio/video clips that make it effective for 3-D and artificial world. Forty-two percent of the of the sample companies use text, graphics and photographs in their Web pages, and twenty-seven per cent use sound or video, in addition to text, graphics and photographs. Thus, the majority of the samples are making use of multimedia in order to attract users and give potential customers a clearer picture of their products.

Framework of Distributed Virtual Environment System

The theories and techniques of virtual environment system is to extract the semantic details of video clips, scenes, single and multiple images, video sequences, robotic navigation and manipulation control, visual data management, human-machine interactions computer, image-based graphical modelling, etc. The visual data in terms of images and videos can be annotated and retrieved for the distributed system and tele-operation of robotic devices. This system is also useful for virtual figure animation and behaviour control and it deals the scene understanding and image-based modelling. The dynamic environment within distributed framework is used to control the Internet based robots for suppressing the system latency and improving the operation efficiency if time-varying delays take place in the Internet communication. The technical implementation for this framework is used for visualization with the help of various computer languages, such as Java, 3-D graphics APIs, C#, ASP.NET, etc. It is a standard for delivering 3-D rendering on the Internet, just like HTML is a standard for Web pages. This system comes from the 'worldly' imitation that distributed virtual environment system files get their name. The files are called 'worlds' and have *.wrl* extension in the multimedia environment. The sound objects are used with controllable attenuation to make the system efficient. It describes irregular ground terrains and extrusion objects for advanced, through compact modelling. For example, a more powerful background coloring and panorama system and a fog system allows the creation of virtual world underwater and cloudy environments. This system is also able to use the MPEG video as a texture map. It interacts with system collision detection gives the user a sense of substance as they move in the world in which touch sensors allow reactions to a users deliberate actions and the proximity sensors allow reactions to a user's not so deliberate actions, whereas visibility sensors allow conservation of resources.

The interpolators provide engines to implement animation. Scripting in JavaScript allows from making programming part from simple logic device to fully blown analytical

NOTES

engines providing a wealth of complexity. The prototypes extend the existing variety of object types with efficient reuse and simple scene graph structure. A navigation information object provides the browser software with details of the speed and nature of the users movements in the world. This system basically uses two factors in the programming part if you use and they are known as nodes and fields. Nodes basically represent a world that is made up of nodes which are types of objects, whereas fields describe properties of a node. For example, if you define Box, Cylinder, Cone, Sphere they would be initialized in the following way:

```
geometry Box {size 5.5 3.75 1.0}
    //Box
geometry Cylinder {radius 0.5 height 10 top FALSE}
    // Cylinder
geometry Cone {bottomRadius 5 height 10 side TRUE
bottom FALSE}
    // Cone
geometry Sphere { radius 10,000,000}
    // Sphere
```

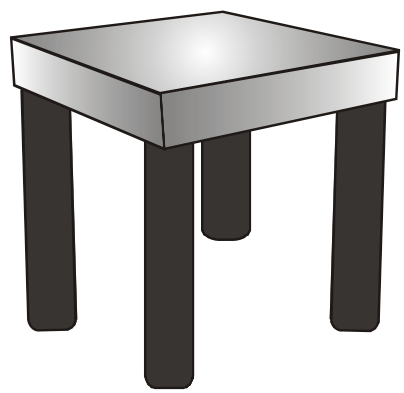


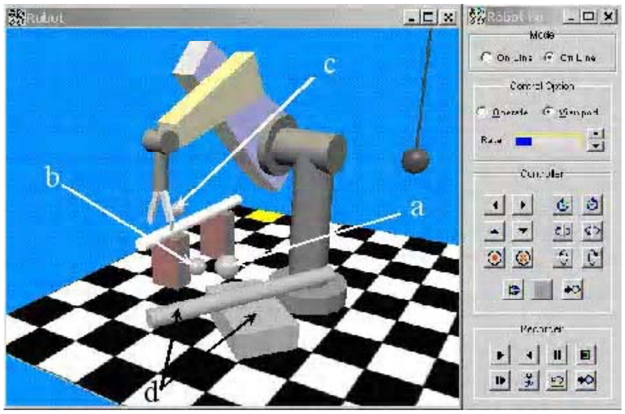
Fig. 6.7 *Shows the 3-D Image of Table in Virtual World*

The techniques used for distributed virtual environment system is computer graphics, imaging, vision, artificial intelligence and networking. The synchronized virtual environment is used with physical counterpart. The online virtual environment is composed and edited for collision detection and path planning so that six Degree-Of-Freedoms (DOF) manipulators, which are useful to implement the interface for system control, remote operation and task management. There are two ways to implement the virtual tour in distributed virtual environment system:

- Using JavaScript to control the viewer's camera position.
- Using VRML sensors, in particular the time sensor.

The tour basically is implemented by using the time sensor approach in which an array of tour positions is defined together with a way to interpolate between them. A touch sensor determines when the timer is started and each tick of the timer is routed to the position interpolator which sets a new orientation and translation. The following screen shows the virtual chess played online by robot across the Internet so that various users can interact with this distributed virtual environment.

NOTES



Screen above shows the virtual environment in which robot plays its chance with the help of artificial intelligence. The chip is set with programming part. It is modelled with system simulation

Examples of DVSE: DVSE presents a computer-generated virtual world to support grid-enabled service oriented framework for facilitating the construction of scalable system on computing grids. For this, a service component called 'gamelet' is proposed, whose distinctive mark is its high mobility for supporting dynamic load sharing of multimedia applications. Gamelet is a mobile service component, which is responsible for processing the work load introduced by a partitioned virtual environment. For example, in a virtual based environment, one or various rooms can form one partition. In a session-based virtual environment, overlapping partition technique is used to support the smooth visual interactions among participants across multiple partition.

Some of the popular examples used for distributed virtual environment system are described as follows:

CAVE Automatic Virtual Environment: This virtual environment system is projection-based virtual reality system. In fact, it is ten feet cubed room. Stereoscopic images are rear-projected onto the walls creating an illusion that 3-D objects exist with the user in the room (see Figure 6.8). The user wears liquid crystal shutter glasses to resolve the stereoscopic imagery. An electromagnetic tracking sensor attached to the glasses allows the CAVE system to restablish the location and orientation of the user's head. This information is used to render the imagery from the user's point of vision. The user can physically walk around an object that appears to exist in 3-D in the middle of the CAVE. The user holds a wand which is also tracked and has a joystick and three buttons for interaction with the virtual environment. The buttons can be used to change modes grab and virtual objects. VR applications displayed on the CAVE can be linked over high-speed networks.

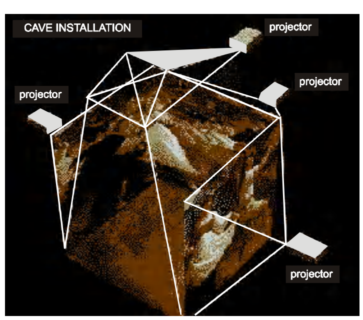


Fig. 6.8 *CAVE Installation*

NOTES

Access Grid Augmented Virtual Environment (AGAVE™): The AGAVE™ display system is a passive stereographic projection system through which audiences will view the immersive content using 3-D movie glasses (see Figure 6.9) . The display system consists of a pull-down polarization-preserving silver screen, two LCD projectors with linear or circular polarizer's in front of each lens, a dual processor Linux driven PC with a high-end graphics card capable of dual display output. To support 3-D tracking and interaction, Ascension Technology's Flock of Birds (Extended Range Transmitter version) and EVL's Wanda can be used. Wanda is a wand with an imbedded tracker, joystick and three buttons. It is the standard interaction device used in CAVEs. The overall concept behind AGAVE is to append PC-based graphics workstation to an Access Grid node that can be used to project 3-D stereoscopic computer graphics to allow networked people to immersively share three-dimensional content. VR applications displayed on the AGAVE can be linked over high-speed networks.

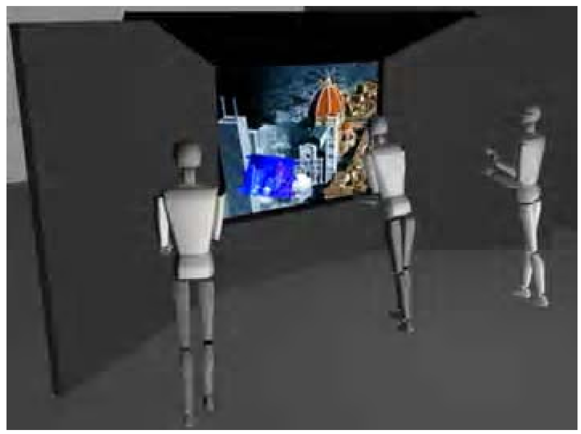


Fig. 6.9 AGAVE™ Environment

The VR application is primarily designed to run in the CAVE™ and on the AGAVE™ to present the distributed virtual environment system. These two popularly environment system can run either locally or through remote networking in both Silicon Graphics, Inc (SGI) and Linux platform. The innovative flexible tools are used for creativity, learning, communication and exchange of knowledge and culture for local and networked visitors.

6.6 VIRTUAL ENVIRONMENT DISPLAYS AND ORIENTATION TRACKING

Virtual environment displays refer to a possibility of applying virtual reality technology. It highlights a synthetic construction using Virtual Reality Modelling Language (VRML) and appropriate simulation technique. The main characteristic of this environment is to include remote environments for the computer-supported cooperative work, entertainment and space exploration, scientific and architectural visualization. This synthetic environment includes two main types of systems known as immersive or non-immersive virtual system. Immersive virtual environment display was explained by Morgan and Zampi as 'an application in terms of quasi-physical experiences'. It is achieved by data gloves and multimedia head mountain display devices whereas non-immersive virtual ED is based on desktop VR and enables screen interface. It creates

> **Virtual environment displays:** A possibility of applying virtual reality technology. It highlights a synthetic construction using Virtual Reality Modelling Language (VRML) and appropriate simulation technique

a feeling of simulated spatial environment. VRML can be universally accessed. It is a hyperlinked and facilitated environment for file formatting for 3D interactive and virtual environments. It represents static and animated objects which are hyperlinked with other media, for example, movies, sound and images etc. The multimedia-based Website is built as a dynamic XHTML code that embeds VRML constructs.

NOTES

The virtual environment is passed through using three main phases which make use of system effectors. They are known as modelling, rendering and run time rendering presentation.

Modelling

The various types of displays use the synthetic construction of objects. For example, Figure 6.10 shows a layout of a building.

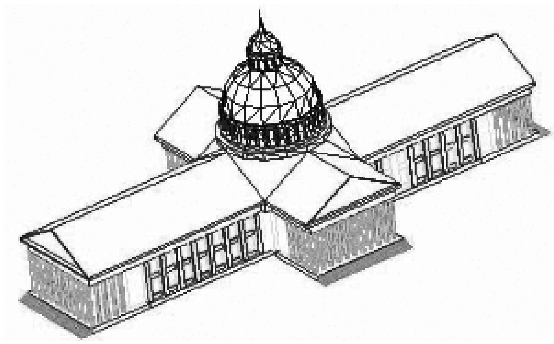


Fig. 6.10 *Construction before Modelling*

Sipes described the modelling process in 1994 for the first time. He also presented loopholes for such kind of a process depending on the information database and confirmation of the computer hardware.

Rendering

This process has been described by Grabowiski and it includes texture maps, lighting parameters etc. Rendering time depends on the quality of rendering, complexity of model and hardware speed (see Figure 6.11). Lindhult chose the same guidelines for inputting texture including quality, versatility, file size and computer speed. He suggested that quality depends on the resolution that is required and essential for presentation in virtual environment displays.

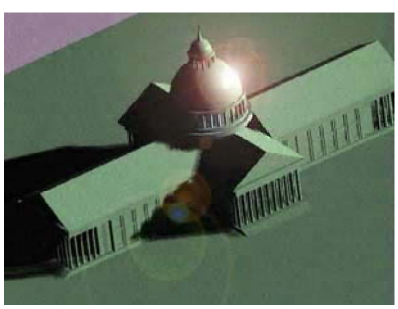


Fig. 6.11 *Construction after Modelling*

Run time Rendering Presentation

It is considered as an attempt to represent the environmental reality and is also known as environmental simulation. The characteristic of environmental simulation is for observer (client) to predict the responses to a real situation. The potential realities

NOTES

contain a designer's imagination along with sound effects. However, it is dependent on the limitations of hardware performance. It is a combination of reality engine and silicon graphics. The VRML is suggested as a future desktop for virtual environment displays (see Figure 6.12).



Fig. 6.12 *Cosmo Player*

Orientation Tracking

The multimedia environment mainly focusses on orientation tracking that provide a complete system to manage media information, such as images, texts, graphics, sounds and image sequences, etc. The stored and manipulated information is used to give virtual environment displays. They are viable in terms of hardware and available for numeric databases and traditional text. Orientation tracking (with reference to virtual environment displays) refers to the process of orienting and assigning the various media information to a circular track, for example, manipulating image and video data includes tasks, such as searching and sorting of visual information . This technique is used for tracking a user-selected object through digital image sequences in multimedia systems. Tracking objects and multimedia information maintain a sequence that is useful for that type of situation where the motion of the object is important. This technique is useful when moving or animated objects are highlighted in a sequence. For example, a particular object moves amongst various similar objects. For this, a tracking tool is used by the multimedia developers. It marks up the moving object so that the next frame waiting in sequence can provide objects orientation to the first frame. The director software package is very useful to set up the frames of animated and moving objects. It also provides a small search area to the objects in each frame of the sequence. It changes the overall illumination between the frames of the sequence. The average intensity of the templates provide arbitrary 2D shapes in an image representing a set of vectors that correspond to the reference point position and each cell denotes the incremented position that corresponds reference point position. This reference point denotes the position of the cell, which is rotated to each vector in the object representation to decide the first frame of the sequence. 3D includes the object's height, weight and length, whereas 4D is added to the accumulator space and object representation vectors can be lengthened and shortened when incrementing the cells. It is associated with the motion-modelling technique in which an object is kept as x and y motion of a test object. At each step, the next position of the object is predicted, whereas the actual position of the object is used to refine the modelling of x and y motion components. The vision-based system follows passive landmarks in orientation tracking that integrates with the hybrid approach in sensor technology. This mechanism revolutionizes a close range of image collection and analysis for multimedia applications. It creates a new way of processing and handling the audio and video datasets by

providing the real-time 3D techniques and approaches. The orientation tracking approach is used in ground level motion imagery using 3D virtual model as control information, such as motion imagery orientation using virtual reality.

Orientation tracking uses following types of VR systems:

- The tracker involved in orientation tracking describes the object orientation to control the process and the signal is detected by the sensor.
- It also controls the unit, which is involved in the process of the signal and sends information to the CPU.
- The signals emit the process to sensors including electromagnetic signals, optical signals, mechanical signals and acoustic signals that produces the virtual environment displays.
- The trackers then calculate the time for this system in which the sensors involved hit the sound for virtual reality applications.

A real 3D position and orientation tracking system are used with wearable devices. This system is combined with wearable reality systems. The system produced is combined with infrared markers with a head-mounted stereo camera that decides the user position and orientation sensor to integrate the signals acquired by multiple sensors. It maintains the accuracy of the system to allow virtual objects and annotations to be overlaid on real scenes through the head-mounted display.

6.7 VISUALLY COUPLED SYSTEM REQUIREMENTS

A **Visually Coupled System (VCS)** is defined as the subsystem of the multimedia environment, which was developed by Birt and Task in 1973. The three components of VCS are a head or helmet-mounted (head-directed) visual display, a means of tracking head of eye-painting director and a source of visual information. In this mechanism, an operator looks in a particular direction where head or eye tracker decides the direction. The visual information source determines the viewing of appropriate images produced by the operator. The visual information source determines viewing of appropriate images produced by the operator. This information is associated with the physical imaging sensor that produces the computer-generated image in the virtual environment system. This mechanism is a subsystem of VCS which is present in both the real and virtual world. It is also referred to as BOOM display. It represents an advances feature of the man-machine interface. This mechanism is designed to check the virtual reality system problem and is implemented as a modular system. It integrates the visual skill of the operator along with the control of the machines. In this, the operator finds track of an object. In essence, the operator looks on the target and sensors or weapons point to the target. The two functions performed in this system are display feedback function and line-of-sight sensing or control function. These functions are used separately as per requirement. It represents a special subsystem which is integrated with the operator's vision. It requires visual coupling applications for aircraft, fire control, weapon delivery, navigation, etc. so that multimedia applications are prepared to make the people aware of those applications. The requirements of the visually coupled system with reference to multimedia applications and the virtual environment are as follows:

- It needs brightness contrast ratios between display of the applications and the outside world.

NOTES

Visually Coupled System (VCS): Defined as the subsystem of the multimedia environment

- It requires the best quality image and display capability under vibration for objects.
- It needs the track orientation technique in display scaling and display format for each specific application.
- It also needs a see-through model presenting themes for multimedia applications which involve a mechanism of the CRT face plate to the observer's eye.
- This mechanism also needs ambient illumination and dual-resolution fields of view with a zoom capability and low resolution for multimedia applications. For example, high-magnification sensors and high-power optics of an object make applications more feasible.
- It requires head-referenced viewing.
- It requires realistic interaction with virtual objects, which allow manipulation,

NOTES

6.8 INTELLIGENT VIRTUAL REALITY SOFTWARE SYSTEMS

The intelligent Virtual Reality Software System (VRSS) architecture implements the notion of an intelligent virtual environment, in which alternative reality can be defined through a symbolic description of the virtual world's behaviour. This software architecture is based on an integration layer, which consists in an event-based system, relating the visualization engine to the behavioural layer. For this, you can use a state-of-the-art game engine for the visualization engine. Game engines provide sophisticated visualization features and most importantly constitute a software development environment in which resources and components are integrated. The intelligent VR software system is the multiple domains of virtual trust between users, resources (hardware as well as software) and services. This system is characterized by scalability distribution, transience and flexible–dynamic nature. The two prime categories for intelligent virtual reality software system are toolkits and authoring systems. The toolkits represent the programming libraries residing in C or C++ which provide virtual functions and help to create VRSS. The authoring system is a complete program for graphical interfaces to create a virtual world without resorting to the detailed programs. It uses scripting language to describe complex actions for defining the virtual reality software system. The benefits for intelligent virtual reality software systems are as follows:

- They help in better presentation of real-life applications and exploitation of resources that present a complete virtual system.
- They allow effective utilization of features of multimedia applications, such as audio/video files, 3D effect, etc. that collectively produce a VRSS.
- It provides a pool of multiple network-storage devices that enable a cross-organizational sharing of data and information.

The intelligent virtual reality software system includes multimedia software, VRML etc. whereas, the multimedia software mainly uses Adobe Authorware, Adobe Director, Winamp, etc. Table 6.2 shows the multimedia software and their functions:

Check Your Progress

1. Define the term intelligent multimedia system.
2. What is virtual reality?
3. What does BOOM system stands for?
4. What do you mean by VROS?
5. What is the use of stereoscopic viewing?
6. Define the term VRML.
7. Which technology supports distributed virtual environment system?
8. What does virtual environment displays refer?
9. On what factors does rendering time depend?
10. Define the term orientation tracking.
11. Who developed Visually Coupled System?

Table 6.2 *Software Used in Multimedia and their Functions*

Software	Function
Adobe Macromedia Authorware	This software uses visual rich-media solution to create online virtual applications.
Chrome Library	This includes free models and samples in the Chrome Library set of 3D behaviors contains 60 Shockwave 3D behaviors that have been designed and developed to meet the needs of 3D graphic designers and developers who are beginners when it comes to Macromedia Director or real-time 3D. It is also useful for developers who want to quickly set up their scenes. The Chrome Library was created for this purpose to make web 3D in Director easier and the inner wonders of the 3D engine more readily available for multimedia applications.
Lingo	Lingo provides a completely integrated package that includes a powerful language for expressing optimization models, a full-featured environment for building and editing problems and a set of fast built-in solvers. This software provides animation, sound, scripting, shockwave 3D for Lingo language.
Shockwave Demo	This software Audio data represents sound object or a mixer can be accessed, processed and passed back to the audio engine and the audio engine will play the processed audio chunk.
Adobe Director	This software contains rich audio capabilities with 5.1 channel surround sound, real-time mixing and digital Signal Processor (DSP) that is used to filter for MP4, H.264, FLV, and F4V video formats and video streaming. The powerful 3D supports including Google SketchUp importer and Byte-Array datatype.
VRML tools	These tools include Chisel (Java-based optimization tool), Seamless3D (is an open source 3D modeling software that exports and imports VRML files), SwirlX3D editor (visual authoring system for VRML which provides full VRML editing with the ease of use of a graphical environment), TriVista technology (provides 3D ImageScene and 3D ImageCube tools for displaying photographs), VizUp (optimize 3D models with polygon reduction tools supporting VRML and other 3D formats.
Digital ArtForms	This software is used in 3D interaction, immersion, and real time SmartScene.
Internet Explorer or Mozilla Firefox	These two executable files provide lease line connection to high broadband services for Internet connection.
Adobe FantaMorph	This powerful software provides fantastic image to morph and warp the movie in real-time 3D. it accelerates hardware and renders the speed easily at the rate of Frame Per Second (FPS). This high speed makes the final play in real time without exporting the file from other application. This software supports BMP, JPEG, PNG, TGA, PCX, TIFF and 32-bit alpha formats. The exported image sequences supported by this software are Adobe Flash Player, animated GIF, image sequence etc. The image can be cropped, rotated, flipped and adjusted source images.

Figure 6.13 shows a videotaped cup on a rotating platform. It presents a scene structure for the virtual intelligent system that is relative to the distance of objects along the tagged frames in the scene. In many ways, the videotaped image is based on techniques that consist of a set of images for the virtual scene and their corresponding depth. When the depth of every point in an image is known, the image can be re-rendered from any nearby point-of-view by projecting the pixels of the image to their 3D locations and re-projecting them onto a new image plane. This cup presents a virtual effect which corresponds to this particular frame of the total system by pixel merits, in which each frame is logically linked to move the graphic image of cup. It is videotaped through programs.

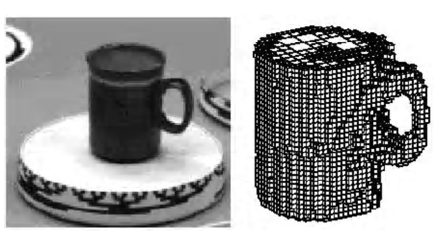


Fig. 6.13 *A Videotaped Cup*

NOTES

Check Your Progress

12. Fill in the blanks with appropriate words:
    a. A ___________ is an object lens system which permits users to construct agents performing a variety of user-defined actions.
    b. In ___________technique, the illusion of immersion is created by projecting stereo images on the walls and floor of a room-sized cube.
    c. __________ is a standard tool for the modeling of three-dimensional virtual environments that are functional and interactive.
    d. In DVE system, each user is represented as an entity called __________ whose state is controlled by the input commands.

VRSS utilizes the input and output channels to interact with virtual environments. VR applications are generic software systems that provides user toolkits and APIs, such as OpenGL, ray-tracing systems etc. The characteristics of intelligent VRSS are as follows:

NOTES

- It provides special display technologies, such as LCD shutter glasses, head-mounted displays, dome projectors, hand-gloves, joysticks and compatible game-pads etc.
- This system supports 3D sound effects using VRML tools and 3D formats for distributed and networked environment so that the VR software becomes intelligent and presentable.
- This system can connect with video monitors to a DVD player, video game consoles and virtual reality applications.
- It provides a new form of communication between machine and man. The basic idea behind it to present an artificial computer world in which users can interact by using a data helmet and gloves. It provides an interface for robot control and automation components to the operator.
- It is suited for universal man–machine interface to control and supervise the class of automation components used in the virtual software system.

Check Your Progress

13. State whether the following statements are true or false:
    a. The audio data is recorded in real time by passing to the system unit. For this, RS232 COM port is used to process the data smoothly.
    b. The VROS or Virtual Reality Operating Systems provides an architectural framework that allows an extensible collection of only interfaces but not protocols and network security management systems.
    c. The virtual reality data maintains the addresses of artificial world and spatial databases so that distributed virtual environment can structure and store the data and control to the artificial world.
    d. Rendering time depends on the quality of rendering, complexity of model and hardware speed.

6.9 SUMMARY

In this unit, you have learn that:

- Intelligent multimedia system provides an advanced technology for listening music and watching movies.
- Virtual reality is basically a way of simulating or replicating an environment and giving the user a sense of being present there, taking control and personally interacting with the environment with his/her own body.
- Virtual reality operating system consists of several software subsystems and requires extra time to reconfigure.
- VROS or Virtual Reality Operating System avoids hardwired configurations because a participant in the virtual world is free to engage in all the resources of the operating system.
- The distributed virtual environment system is an Internet based multi-user virtual reality system in which participants are able to navigate the 3-D virtual world and also able to work with various applications and other users.
- Virtual environment displays refer to a possibility of applying virtual reality technology. It highlights a synthetic construction using VRML or Virtual Reality Modeling and appropriate simulation technique.
- A visually coupled system is defined as the subsystem of the multimedia environment, which was developed by Birt and Task in 1973.
- The three components of VCS are a head or helmet-mounted (head-directed) visual display, a means of tracking head of eye-painting director and a source of visual information.
- VRSS or Virtual Reality Software System implements the notion of an intelligent virtual environment, in which alternative reality can be defined through a symbolic description of the virtual world's behaviour.

6.10 ANSWERS TO 'CHECK YOUR PROGRESS'

NOTES

1. Intelligent multimedia system provides an advanced technology for listening music and watching movies. This system allows users to avail the microphone facility as an analog input to sing and speak.
2. Virtual reality is basically a way of simulating or replicating an environment and giving the user a sense of being present there, taking control and personally interacting with the environment with his/her own body.
3. BOOM stands for Binocular Omni-Orientation Monitor.
4. Virtual Reality Operating System (VROS) consists of several software subsystems and requires extra time to reconfigure.
5. Stereoscopic viewing enhances the perception of depth and the sense of space.
6. VRML or Virtual Reality Modeling Language is the standard tool for the modelling of three-dimensional virtual environments that are functional and interactive.
7. VoIP technology supports distributed virtual environment system.
8. Virtual environment displays refer to a possibility of applying virtual reality technology. It highlights a synthetic construction using Virtual Reality Modeling Language (VRML) and appropriate simulation technique.
9. Rendering time depends on the quality of rendering, complexity of model and hardware speed.
10. Orientation tracking refers to the process of orienting an assigning the various media information to a circular track, for example, manipulating image and video data includes tasks, such as searching and sorting of visual information.
11. VCS or Virtually Coupled System was developed by Birt and Task in 1973.
12. (a) Quintessential, (b) CAVE, (c) VRML, (d) Avatar.
13. (a) True, (b) False, (c) True, (d) True.

6.11 QUESTIONS AND EXERCISES

Short-Answer Questions

1. List the characteristics of intelligent multimedia system.
2. What is the use of BOOM system?
3. Write short note on VRML.
4. What do you mean by run time rendering presentation?
5. What are requirements for visually coupled system?

Long-Answer Questions

1. Discuss the various applications of virtual reality with the help of suitable illustrations
2. Explain the significance of DVE system.
3. What do you mean by orientation tracking? Explain with the help of example
4. How to implement the virtual tour in distributed virtual environment system? Describe the steps involved.
5. Write in detail about intelligent virtual reality software. Discuss its significance.

MODEL QUESTION PAPER
MBA Degree Examination
Multimedia and Applications

Time: 3 Hours **Maximum: 100 Marks**

PART A (5 × 8 = 40 marks)

Answer any FIVE of the following:

1. Trace the evolution of multimedia and its various components.
2. Enlist the various types of multimedia developmental tools.
3. Differentiate the types of media and write their areas of application.
4. What is the significance of speech recognition in multimedia?
5. Explain the various standards for image compression.
6. What do you understand by the term 'object oriented multimedia'? What hardware support is required for object oriented multimedia?
7. Compare the various types of CD technologies giving their significant features.
8. Explain the significance of virtual reality in multimedia.

PART B (4 × 15 = 60 marks)

Answer any FOUR of the following:

9. Explain the major features and working of the multimedia. Why multimedia is used to develop interactive Websites?
10. Classify the various media types. Explain the scope and importance of each media type in developing multimedia applications.
11. Explain the process of compression in multimedia. Elaborate on video and image compression standards.
12. Why object oriented multimedia is gaining importance? Justify your answer giving examples.
13. Explain the various types of media technology and their architectural structure typically used in multimedia
14. Discuss the significance and applications of intelligent multimedia systems.

Compulsory

15. Why virtual reality term is applied to computer simulated environments that can simulate physical presence in places in the real world as well as in imaginary worlds?